youth futures FOUNDATION

Theory-informed implementation and process evaluation: A rapid review of guidance

Dr Stephanie Smith, Dr Gayatri Kembhavi-Tam, Dr Paula Verdugo, Amy Hall, Jamie Rowland, Jane Lewis (CEI)

Ediane Santana de Lima, Cristina Preece, Finlay Green, Dr Julie Harris (Dartington Service Design Lab)

Centre for Evidence and Implementation

DARTINGTON SERVICE DESIGN LAB

Table of Contents

About the Authors

Youth Futures Foundation commissioned the Centre for Evidence and Implementation (CEI) and Dartington Service Design Lab to undertake a rapid evidence review on theory-informed implementation and process evaluation (TIIPE), with a focus on youth employment where possible. The review addresses a gap in existing evaluation guidance by examining how programme theory and implementation theory can be used together to strengthen understanding of what works, for whom, and under what conditions, and to inform the design of high-quality evaluations and delivery in real-world contexts.

Centre for Evidence and Implementation
Dr Stephanie Smith (Associate Director)
Dr Gayatri Kembhavi-Tam (Associate Director)
Jane Lewis (Associate Director)
Paula Verdugo Phd (Advisor)
Amy Hall (Advisor)
Jamie Rowland (Advisor)

Dartington Service Design Lab
Ediane Santana de Lima (Senior Research Lead)
Cristina Preece (Research and Design Officer)
Finlay Green (Strategic Lead for Evaluation)
Dr Julie Harris (Director of Research)

Youth Futures Foundation (funders and project oversight)
Jane Mackey (Evidence and Evaluation Manager)
Hannah Murphy (Head of Evaluation)

Suggested citation
Smith, S.*, Santana de Lima, E.*, Kembhavi-Tam, G., Preece, C., Verdugo, P. Hall, A. Rowland, J., Green, F., Harris, J.†, Lewis J.† (2026). *Theory-informed implementation and process evaluation: a rapid review of guidance*. Youth Futures Foundation: London.

* *Joint first authors*
† *Joint senior authors*

Acknowledgements

The Centre for Evidence and Implementation (CEI) is an independent, not-for-profit organisation dedicated to improving outcomes for children, families and communities by closing the gap between what we know from research and what we do in practice. With teams based in the UK, Australia, Europe and Asia, CEI works with policymakers, practitioners, organisational leaders and funders to translate evidence into action, support the effective implementation of policies and programmes, and generate robust evidence on what works, for whom, and in what contexts.

Dartington Service Design Lab is an independent research and design charity that harnesses experience, cutting edge evidence and design to tackle the root causes and conditions that sustain the most pressing challenges children and young people face today. Collaborating closely with young people, communities and public service leaders, Dartington's team works to design, test and embed changes that improve outcomes for children and young people. Dartington Service Design Lab brings a strong tradition of interdisciplinary working and a practical focus on improving services and systems for children and their families.

This report was produced through a close collaboration between CEI and Dartington Service Design Lab. We are grateful to colleagues across both organisations whose expertise, commitment and thoughtful engagement made a significant contribution to its development.

We extend our thanks to Youth Futures for commissioning the project. We are particularly grateful to Jane Mackey for her collaborative approach throughout and for the wise support and advice she provided at each stage. We also thank Hannah Murphy for her role in shaping the final guidance and for steering the project through to completion.

We would also like to acknowledge and thank members of YFF's CAT-BEE advisory group, the TIIPE user advisory group, and CEI's Methods Lab, whose insights and reflections were invaluable in grounding the work in the realities of implementation and evaluation practice.

Finally, we wish to offer special thanks and recognition to the late Dr Jane Colechin, whose original idea and early vision sparked this commission. Jane's enthusiasm, intellectual generosity and deep commitment to strengthening the relationship between theory and evaluation shaped the foundations of this work.

Contact:

Get in touch

Youth Futures Foundation

Fivefields,
8-10 Grosvenor Gardens,
London,
SW1W 0DH

Email: hannah.murphy@youthfuturesfoundation.org

www.youthfuturesfoundation.org

1. Introduction

1.1. Background and rationale

Evidence of effectiveness is not enough, on its own, to inform policy and practice. Robust programme evaluation also involves understanding how a programme creates change (or does not), whether programmes are being implemented as intended, and what it takes for them to be implemented well. This is fundamental knowledge for decision-makers to determine whether a programme is right for their context, and how to make it work there. Several of the What Works Centres have produced guidance for evaluators on how to conduct implementation and process evaluations (IPEs). However, there is less guidance on the role of theory in evaluation and how to undertake IPEs that are informed by both programme theory and implementation theory (that is, theory-informed implementation and process evaluation, (TIIPE)).

To address this gap, Youth Futures Foundation (Youth Futures) commissioned the Centre for Evidence and Implementation (CEI) and Dartington Service Design Lab (DSDL) to undertake a rapid review of evidence on TIIPE, where possible with a focus on youth employment. This review will be used to create cutting-edge cross-sector guidance that will support evaluators to utilise rigorous methods to develop, apply, evaluate, and refine programme theory, and to use implementation theory in the design and delivery of high-quality IPEs. This report describes the approach to the first stage of work - the rapid evidence review - and synthesises the evidence reviewed.

When discussing the 'theory' in TIIPE, we mean both programme theory and implementation theory – with each playing a distinct but complementary role. Programme theory articulates the hypothesised causal pathways through which activities are expected to lead to desired outcomes, including describing the underlying assumptions that link inputs, mechanisms, and impacts. Implementation theory, by contrast, outlines the contextual factors and conditions required for a programme to be delivered as intended – and therefore for the programme theory to have a realistic chance of playing out in practice. Both theories are therefore essential to uncovering '***what works, for whom, under what conditions***'.

In this review, we distinguish clearly between programme theory and implementation theory and treat both as integral to robust evaluation. Programme theory helps evaluators to understand what is supposed to happen and why, while implementation theory helps explain whether and how delivery unfolds within real-world systems. This distinction has deep roots in the evaluation literature. Suchman (1967) was the first to differentiate implementation failure (e.g., the programme was not delivered as

intended), from theory failure (e.g., the programme was delivered as intended, but the underlying causal logic was flawed). Later, Weiss, (1995; 1997) reinforced and elaborated this distinction, emphasising the need to examine both the plausibility of the programme model and the conditions required for successful delivery. Drawing on this lineage, our review treats programme theory and implementation theory as complementary explanatory lenses – each essential for understanding variation in outcomes across contexts and populations.

By examining the existing literature through this dual-theory lens, the review aims to identify how programmes conceptualise change, how they operationalise delivery, and how implementation conditions shape the likelihood of achieving intended impacts. This framing supports stronger interpretation of evaluation findings, and provides a foundation for theory-informed design, adaptation, and scale-up.

The review incorporates academic texts on theory in evaluation and other existing guidance documents on conducting IPEs and evaluations of complex interventions, including Humphreys et al.'s (2016a; 2016b) literature synthesis and handbook which have influenced the work of the Education Endowment Fund (EEF) and other What Works Centres. Our starting point was scoping work to develop an a priori set of key topics we expected to cover, based on a review of the coverage of key guidance documents (including Humpreys et al. 2016a and 2016b, Skivington et al., 2021's MRC guidance on evaluating complex interventions, Bickett et al. (2020) and supplements, the ImpRes tool (King's Health Partners, 2019) and the guidance produced by other What Works Centres, including the Early Intervention Foundation, EEF, Youth Endowment Fund (YEF) and Transforming Access and Student Outcomes in Higher Education (TASO).

The goal of this review was to build on existing guidance in two substantial ways. First, we aimed to extend and update coverage of the development, testing and refining of programme theories of change, and how this could be incorporated into different stages of evaluation. Second, we aimed to explore how implementation theory is applied in evaluation and the use of key implementation constructs in evaluations: implementation outcomes (e.g. Proctor et al., 2011), implementation determinants (e.g. Damschroder et al., 2022; Moullin et al., 2019) and implementation strategies (e.g. Waltz et al., 2015).

1.2. Research questions

With this context in mind, this rapid evidence review sought to answer four research questions:

- **RQ1:** How is TIIPE defined and its function in different evaluation contexts described?

- **RQ2:** How should evaluators approach the development, testing and refinement of programme theory and implementation theory?
- **RQ3:** How should evaluators design evaluations to test programme theory and implementation theory, at different stages of evaluation and in different contexts?
- **RQ4**: How has TIIPE been undertaken in evaluations of interventions to support the employment of marginalised young people?

1.3. Overview of review approach

The stages of this review included scoping of the search terms and approach, protocol development, literature search, screening, extraction, synthesis and production of the report. The scoping phase began in March 2024, and the protocol was published in May 2024 on Open Science Framework[1]. The search and screening were conducted between May and June 2024, extraction took place between July -September 2024, synthesis and report writing between October – December 2024 and then finalising the report in December 2025.

We took a pragmatic and progressive approach to identifying papers for this review. There is an extensive literature on programme theory and implementation theory, and therefore it was challenging to employ a conventional systematic review strategy within the timeframe and resources available. As outlined further below, our approach included searching within academic databases with tight search terms and parameters, a targeted website search for grey literature, as well as consultation with experts in the field, and exploration of Youth Futures' published IPEs and Evidence Gap Map, with additional texts included to fill gaps that emerged in synthesis[2].

1.4. Overview of the report

In the next section (Chapter 2), we explain the search strategy, eligibility criteria, and our approach to screening and extraction. Chapters 3 to 6 focus on RQs 1-3. Chapter 3 reviews key definitions and concepts relating to TIIPE adopted in the reviewed studies. Chapter 4 focuses on how to develop, test and refine theory in TIIPE, Chapter 5 looks specifically at equity in TIIPE, and Chapter 6 discusses guidance on how the focus of TIIPE changes at different stages of evaluation. Chapter 7 then turns to RQ 4, providing an

[1] https://osf.io/e2crk
[2] These included specific papers on hybrid designs, and papers providing more detail on examining the role of context and mechanisms.

overview of how theory was incorporated in the selected individual evaluations of initiatives to support marginalised young people into work. Finally, Chapter 8 provides a discussion of the evidence and of the strengths and limitations of the review.

2. Review methods

In this section we first describe the approach taken to identify literature to address RQs 1-3, and then (below) the approach taken to identify evaluation studies reviewed for RQ 4.

2.1 Search strategy

2.1.1. Academic databases

Three academic databases were searched – SCOPUS, PsycInfo, and PubMed, for relevant articles published in English and since January 2000. The final search was completed on 26th April 2024.

2.1.2. Grey literature

Targeted website searches were conducted for grey literature. This list of sites was selected based on expertise within the study team and initial project scoping for relevant publications. The websites searched were:

- Youth Endowment Fund - https://youthendowmentfund.org.uk/
- Early Intervention Foundation - https://www.eif.org.uk/ (archived)
- Education Endowment Foundation - https://educationendowmentfoundation.org.uk/
- Implementation Science Exchange - https://impsci.tracs.unc.edu/
- National Institute for Health Research - https://www.ncbi.nlm.nih.gov/
- UK Research and Innovation - https://www.ukri.org/publications/
- Medical Research Council - https://www.ukri.org/councils/mrc/
- Gov.uk, HM Treasury - https://www.gov.uk/government/publications/
- MRC/CSO Social and Public Health Sciences Unit, University of Glasgow - https://eprints.gla.ac.uk/
- Treasury Board of Canada Secretariat - https://www.canada.ca/en/treasury-board-secretariat/
- International Initiative for Impact Evaluation - https://www.3ieimpact.org/
- Portugal 2020 - https://portugal2020.pt/

- Hivos ToC Learning Group - https://hivos.org/
- Governance and Social Development Resource Centre - http://www.gsdrc.org/
- Government Outcomes Lab - https://golab.bsg.ox.ac.uk/knowledge-bank/
- Polish Agency for Enterprise Development - https://en.parp.gov.pl/
- TASO - https://taso.org.uk/evidence/evaluation-guidance-resources/

2.1.3. Expert consultation

We also took advice from Youth Futures' Complex and Theory-Based evaluation advisory group who reviewed the a priori list of the key topics we expected to cover (see above) and were asked to provide feedback on the content and key texts they expected to see discussed. The experts advised to differentiate our guidance from previous guidance texts by ensuring a central role for implementation science and its application to evaluation, as well as using the opportunity of the review to increase awareness of implementation theory more broadly.

2.2. Search terms

The search terms used in the academic database searches included three layers – texts needed to include 1) **guidance** (i.e. provide practical advice to evaluators on aspects of evaluation) on 2) **theory** (implementation and/or programme theory), in relation to 3) **research or evaluation**. We prioritised texts providing guidance for evaluators on aspects of TIIPE, including reviews. For this part of the review, we excluded texts that were individual evaluations studies, as well as books or book chapters. The search covered the broad policy area of human services, and no geographic restrictions were set.

The overall search strategy and search terms were tested and refined in a two-step process. Initially, the drafted search strategy was reviewed by the CEI Methods Lab, a group of internal and external methods experts who provided rapid guidance on key methodological issues. The review team then established a list of 15 target publications (i.e. texts known to the team that were seen as key texts to include in the review) against which the search strategy was piloted and adjusted. Dummy searches were carried out to ensure that the proposed search terms returned most of these known texts.

2.3. Inclusion and exclusion criteria

The inclusion and exclusion criteria used to identify eligible papers is outlined in Table 1.

Table 1: Inclusion and exclusion criteria

DOMAIN	INCLUSION CRITERIA	EXCLUSION CRITERIA
TYPE OF PAPER	• Reviews (broadly defined – including systematic reviews, meta-analyses, scoping reviews, umbrella reviews, rapid reviews, narrative reviews) • Guidance (including toolkits, guides, frameworks, handbooks)	• Individual evaluation studies • Books or book chapters
TYPE OF THEORY	• The review or guidance document specifically focuses on theory testing within an evaluation context using either programme theory and/or implementation theory (which includes implementation theories, models, and frameworks)	• Any other type of theory aside from programme or implementation theory, e.g., systems theory, behavioural theory
POLICY AREA	• Restrict to human services, e.g., health, social care, employment, education.	• Non-human studies, e.g., computer sciences, engineering, environmental sciences
GEOGRAPHIC COVERAGE	• All geographies to be included	• None
LANGUAGE	• Papers published in English	• Papers published in other languages
PUBLICATION DATE	• Papers published since January 2000 to present day	• Papers published before the year 2000

2.4. Screening

All five reviewers attended a screening briefing session in which they screened 15 random papers, in pairs, to ensure consistency in decision making. Screening was then conducted in two stages. In the first stage, titles and abstracts were screened using a prioritised set of inclusion criteria (e.g. reference to theory or evaluation). In the second stage, full texts were screened based on the full inclusion and exclusion criteria.

As this was a rapid review, each paper was reviewed by one reviewer at the title and abstract screening stage. Ambiguous cases were put through to the full text screening stage for more in-depth consideration. At the full text review stage each paper was screened by two reviewers, blind to each other's decision. If the two reviewers disagreed, inclusion to the final stage was decided by team consensus. No bibliographic fields were blinded during the screening process. All included texts were then transferred to the extraction phase.

Prior to extraction, the texts were further reviewed by the full review team and a further 13 texts were added to the extraction pool which had not been identified in the database search. These were identified from a combination of grey literature searching and expert consensus.

2.5. Extraction

Relevant content was extracted onto an extraction template. Extraction was completed in four stages: template development, template piloting, training, and full extraction. The extraction template was developed based on a review of the Research Questions and the topic coverage. Template piloting was conducted by two reviewers, blind to each other's responses, on two of the target publications. Following discussion, the extraction template was adjusted to improve usability, and training provided for the five reviewers involved in data extraction. Data from each paper were extracted by one reviewer. All reviewers are knowledgeable in the study background and in cases where a primary reviewer was uncertain about an extraction field, they discussed it with a second reviewer. More complex questions were addressed and answered within weekly team meetings.

2.6. Synthesis

Extracted data were narratively synthesised to address each research question. Reviewers initially familiarised themselves with the extracted data independently before coming together for three analysis workshops where key themes, patterns and considerations were surfaced and structured.

2.7. Software

All citations identified in the search were imported into the online systematic review application, Covidence. The full screening process (from title and abstract to full text) for these papers was carried out on Covidence to ensure that accurate records were kept of the screening process, including the number of duplicates removed and reasons for exclusion. Texts identified through website searches and through expert advice were managed in Excel. The data extraction template was created in Excel to manage data extraction across the first three research questions.

We also used the Covidence machine learning function, which learns which papers are more likely to be included and sequences them to facilitate faster decision making. Covidence automatically identifies and removes duplicate papers from the screening process, and papers can also be manually marked as duplicates.

2.8. Summary of studies extracted for the rapid review

The PRISMA flow diagram in Figure 1 presents the number of papers screened, assessed for eligibility and extracted. Database searches identified 4,363 records for screening, of which 264 full texts were screened and 37 were included. A further three papers were identified from a grey literature search, giving a total of 40 papers obtained through the formal search and screening process.

During synthesis, the review team identified some gaps and areas where particular concepts were insufficiently represented in the set of included studies. To address these gaps, we incorporated 10 additional papers that were not captured by the search strategy – primarily because key terms relating to the combination of 'guidance', 'theory', or 'research/evaluation' did not appear in their titles or abstracts. These included influential methodological theoretical and methodological texts (Moore et al., 2015; Proctor et al., 2011; Proctor et al., 2023; Waltz et al., 2015; Weiner et al., 2017). We also included further papers focused on specific domains central to TIIPE, including

hybrid trial designs (Curran et al., 2012), implementation research logic models (Czosnek et al., 2022), contexts-mechanisms interactions (Greenhalgh & Manzano, 2021; Lemire et al., 2020) and realist evaluation methods (Renmans & Pleguezuelo, 2023). These 10 texts were identified through expert knowledge and collective discussion within the review team, resulting in a total of 50 papers included in the evidence base for synthesis.

In addition to these, we also draw on a number of further texts that, while not part of the formally screened and extracted set, provide important conceptual foundations or methodological clarification relevant to interpreting the evidence. These supplementary texts are listed in the References section and are marked with an asterisk. As they were not identified through the search strategy or included in extraction, they do not appear in Table 2.

2.9. Characteristics of extracted studies

A summary of characteristics for all extracted texts (n=50) can be found in Table 2. Included texts were published between 2003 and 2024, and 52% were published within the last 5 years. Only two of the included texts were grey literature; the remaining 48 were peer-reviewed articles. The set comprised a mixture of open-access and subscription-only (paywalled) journal articles. Of the included papers, 18 were reviews, 16 were guidance, 14 were frameworks, 1 paper outlined a typology related to study design categorisation, and 1 was a tool for assessing implementation outcomes. Thirty of the texts were drawn from the implementation science literature and 20 were drawn from the programme theory and evaluation literature.

Regarding policy areas, two thirds (n = 32) of the texts were developed for use in healthcare, four in welfare, three in education, two in employment, and nine texts were not developed for a specific policy area.

Eighteen texts were written by lead authors based in North America, 16 in Europe, 12 from international teams, three from Australasia and one from South America. Included texts were written by co-authors located across the world, in Australia, Belgium, Brazil, Canada, China, Denmark, Ethiopia, India, Italy, New Zealand, Norway, South Africa, Sweden, Tanzania, the Netherlands, Uganda, the United Kingdom, the USA, and Zimbabwe.

The target users of most texts were described as being researchers/evaluators (n = 43), four of which were also aimed at providers/implementers, four at policymakers/funders and one at intervention developers. Five texts specifically targeted providers/implementers and two specifically targeted intervention developers.

Figure 1. PRISMA flow diagram

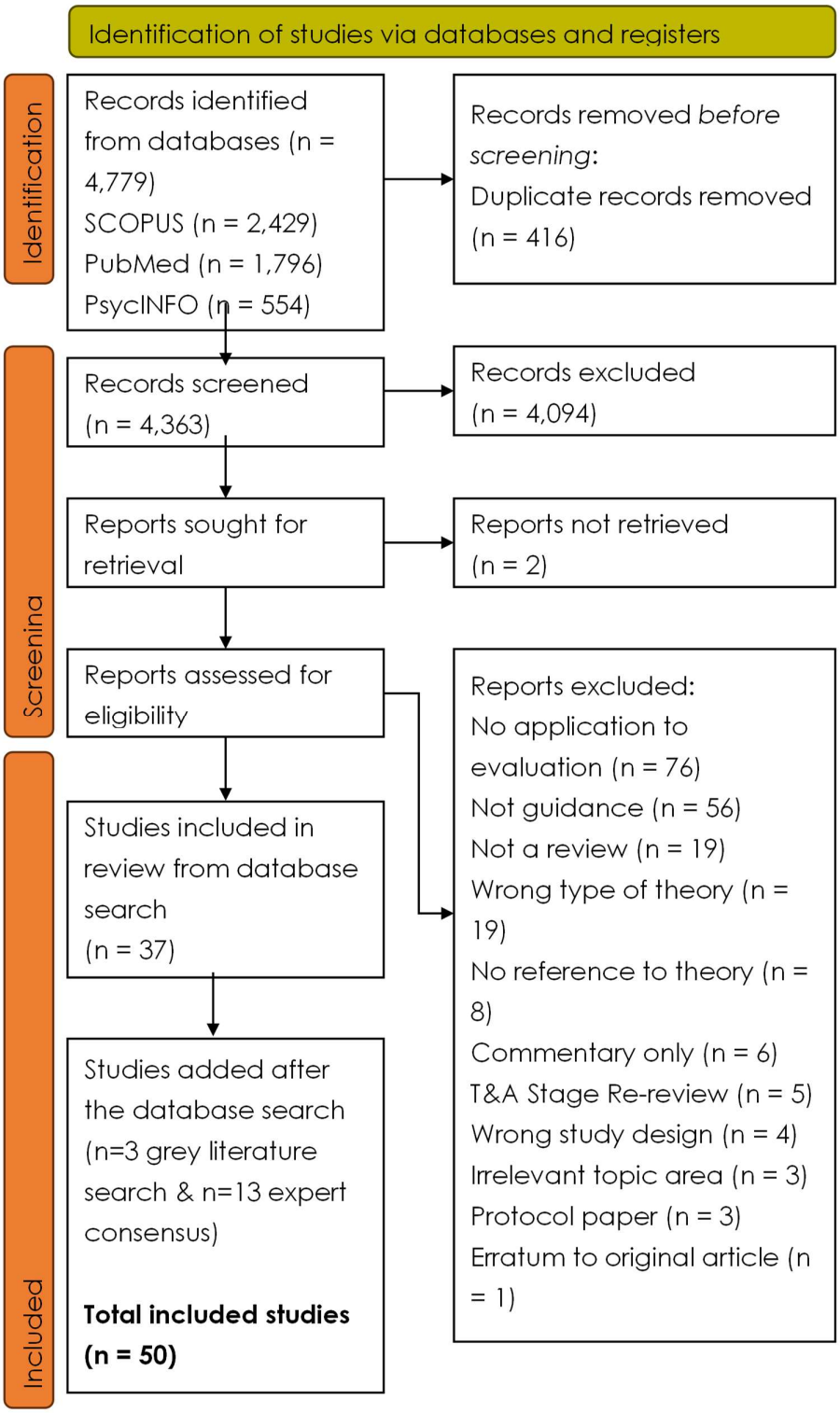


Table 2. Characteristics of included studies

AUTHOR(S)	PUBLICATION TITLE	PUBLICATION YEAR	LOCATION	PAPER TYPE	POLICY AREA	THEORY TYPE
Albers, B. et al.	Implementation frameworks in child, youth and family services: Results from a scoping review	2017	International	Review	Welfare	Implementation
Cargo, M. et al.	Cochrane Qualitative and Implementation Methods Group guidance series - paper 4: methods for assessing evidence on intervention implementation	2018	International	Guidance	Healthcare	Implementation
Carroll, C. et al.	A conceptual framework for implementation fidelity	2007	Europe	Framework	Welfare	Implementation
Craig, P. et al.	Taking account of context in population health intervention research: guidance for producers, users and funders of research	2018	International	Guidance	Healthcare	Programme
Curran, G. et al.	Effectiveness-implementation Hybrid Designs	2012	North America	Typology	Healthcare	Implementation
Czosnek, L. et al.	Developing an implementation research logic model: using a multiple case study design to establish a worked exemplar	2022	Australasia	Framework	Healthcare	Implementation

AUTHOR(S)	PUBLICATION TITLE	PUBLICATION YEAR	LOCATION	PAPER TYPE	POLICY AREA	THEORY TYPE
Damschroder, L. et al.	Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science	2009	North America	Framework	Healthcare	Implementation
Damschroder, L. et al.	The updated Consolidated Framework for Implementation Research based on user feedback	2022	North America	Framework	Healthcare	Implementation
Damschroder, L., et al.	Conceptualizing outcomes for use with the Consolidated Framework for Implementation Research (CFIR): the CFIR Outcomes Addendum	2022	North America	Framework	Healthcare	Implementation
Davies, P., et al.	A systematic review of the use of theory in the design of guideline dissemination and implementation strategies and interpretation of the results of rigorous evaluations	2010	Europe	Review	n/a	Programme
De Silva, M, et al.	Theory of Change: a theory-driven approach to enhance the Medical Research Council's framework for complex interventions	2014	International	Guidance	Healthcare	Programme
Durlak, J. & DuPre, E.	Implementation Matters: A Review of Research on the Influence of Implementation on Program	2008	North America	Review	Healthcare	Implementation

AUTHOR(S)	PUBLICATION TITLE	PUBLICATION YEAR	LOCATION	PAPER TYPE	POLICY AREA	THEORY TYPE
	Outcomes and the Factors Affecting Implementation					
Edmunds, S., et al.	A method for defining the CORE of a psychosocial intervention to guide adaptation in practice: Reciprocal imitation teaching as a case example.	2022	North America	Framework	Healthcare	Programme
Fletcher, A., et al.	Realist complex intervention science: Applying realist principles across all phases of the Medical Research Council framework for developing and evaluating complex interventions.	2016	Europe	Guidance	N/A	Programme
Frye, A. & Hemmer, P.	Program evaluation models and related theories: AMEE Guide No. 67	2012	North America	Guidance	Education	Programme
Greenhalgh, J., & Manzano, A.	Understanding 'context' in realist evaluation and synthesis	2022	Europe	Review	N/A	Programme
Gustafson, P., et al.	A scoping review of equity-focused implementation theories, models and frameworks in healthcare and their application in addressing ethnicity-related health inequities.	2023	International	Review	Healthcare	Implementation

AUTHOR(S)	PUBLICATION TITLE	PUBLICATION YEAR	LOCATION	PAPER TYPE	POLICY AREA	THEORY TYPE
Holtrop, J., et al.	Understanding and applying the RE-AIM framework: Clarifications and resources	2021	North America	Framework	Healthcare	Implementation
Humphrey, N., et al.	Implementation and process evaluation (IPE) for interventions in education settings: A synthesis of the literature	2016	Europe	Review	Education	Implementation
Humphrey, N., et al.	Implementation and process evaluation (IPE) for interventions in education settings: An introductory handbook	2016	Europe	Guidance	Education	Implementation
Kamphuis, C., et al.	Development of a framework to guide research into policies promoting physical activity and healthy diets in the European context: the system-based Policy Evaluation Network (PEN) framework	2022	International	Framework	Healthcare	Programme
Lafferty, C. & Mahoney, C.	A framework for evaluating comprehensive community initiatives.	2003	North America	Framework	Welfare	Programme
Lemire, C., et al.	A comparison of fidelity Implementation Frameworks Used in the Field of Early intervention	2023	North America	Review	Welfare	Implementation

AUTHOR(S)	PUBLICATION TITLE	PUBLICATION YEAR	LOCATION	PAPER TYPE	POLICY AREA	THEORY TYPE
Lemire, S., et al.	What Is This Thing Called a Mechanism? Findings From a Review of Realist Evaluations	2020	North America	Review	N/A	Programme
Lengnick-Hall, R., et al.	Eight characteristics of rigorous multilevel implementation research: a step-by-step guide	2023	North America	Guidance	Healthcare	Implementation
Masterson-Algar, P., et al.	The generation of consensus guidelines for carrying out process evaluations in rehabilitation research.	2018	Europe	Guidance	Healthcare	Programme
May, C., & Finch, T.	Implementing, embedding, and integrating practices: an outline of normalization process theory	2009	International	Guidance	n/a	Implementation
Moore, G., et al.	Process evaluation of complex interventions: Medical Research Council guidance	2015	Europe	Guidance	Healthcare	Programme
Moullin, J., et al.	Ten recommendations for using implementation frameworks in research and practice	2020	International	Guidance	Healthcare	Implementation
Nielsen, K. & Abildgaard, J.	Organizational interventions: A research-based framework for the	2013	Europe	Framework	Employment	Programme

AUTHOR(S)	PUBLICATION TITLE	PUBLICATION YEAR	LOCATION	PAPER TYPE	POLICY AREA	THEORY TYPE
	evaluation of both process and effects					
Nielsen, K., et al.	Quantitative process measures in interventions to improve employees' mental health: A systematic literature review and the IPEF framework	2023	Europe	Review	Employment	Implementation
Nilsen, P.	Making sense of implementation theories, models and frameworks	2015	Europe	Review	Healthcare	Implementation
Ogden, T. & Fixsen, D.	Implementation Science: A Brief Overview and a Look Ahead	2015	International	Review	n/a	Implementation
Pfadenhauer, L., et al.	Making sense of complexity in context and implementation: the Context and Implementation of Complex Interventions (CICI) framework.	2017	Europe	Framework	Healthcare	Programme
Presseau, J., et al.	Action, actor, context, target, time (AACTT): a framework for specifying behaviour.	2019	International	Guidance	Healthcare	Implementation
Proctor, E., et al.	Ten years of implementation outcomes research: a scoping review	2023	North America	Review	Healthcare	Implementation

AUTHOR(S)	PUBLICATION TITLE	PUBLICATION YEAR	LOCATION	PAPER TYPE	POLICY AREA	THEORY TYPE
Proctor, E., et al.	Outcomes for Implementation Research: Conceptual Distinctions, Measurement Challenges, and Research Agenda	2010	North America	Guidance	Healthcare	Implementation
Renmans, D. & Pleguezuelo, V.	Methods in realist evaluation: A mapping review	2023	Europe	Review	N/A	Programme
Ringhofer, L. & Kohlweg, K.	Has the Theory of Change established itself as the better alternative to the Logical Framework Approach in development cooperation programmes?	2019	Europe	Review	n/a	Programme
Romão, D., et al.	Integration of evidence into Theory of Change frameworks in the healthcare sector: A rapid systematic review.	2023	South America	Review	Healthcare	Programme
Sekhon, M., et al.	Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework.	2017	Europe	Review	Healthcare	Implementation
Seward, N., et al.	A guide to systems-level, participatory, theory-informed implementation research in global health.	2021	International	Guidance	Healthcare	Programme

AUTHOR(S)	PUBLICATION TITLE	PUBLICATION YEAR	LOCATION	PAPER TYPE	POLICY AREA	THEORY TYPE
Skivington, K., et al.	A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance	2021	Europe	Guidance	Healthcare	Programme
Vicki, B., et al.	A narrative review of economic constructs in commonly used implementation and scale-up theories, frameworks and models	2020	Australasia	Review	Healthcare	Implementation
Waltz, T., et al.	Use of concept mapping to characterize relationships among implementation strategies and assess their feasibility and importance: results from the Expert Recommendations for Implementing Change (ERIC) study	2015	North America	Framework	Healthcare	Implementation
Wang, Y., et al.	A scoping review of implementation science theories, models, and frameworks — an appraisal of purpose, characteristics, usability, applicability, and testability	2023	International	Review	Healthcare	Implementation
Weiner, B., et al.	Psychometric assessment of three newly developed implementation outcome measures	2017	North America	Assessment tool	Healthcare	Implementation

AUTHOR(S)	PUBLICATION TITLE	PUBLICATION YEAR	LOCATION	PAPER TYPE	POLICY AREA	THEORY TYPE
Wolfenden, L., et al.	Designing and undertaking randomised implementation trials: guide for researchers	2020	Australasia	Guidance	Healthcare	Implementation
Woodland, R. & Mazur, R.	Reclaiming Logic Modeling for Evaluation: A Theory of Action Framework	2024	North America	Framework	n/a	Programme
Woodward, E., et al.	A more practical guide to incorporating health equity domains in implementation determinant frameworks.	2021	North America	Framework	Healthcare	Implementation

2.10. Youth Futures reports and Evidence Gap Map

To address RQ4, we screened the IPEs published by Youth Futures and high and medium quality evaluation studies from Youth Futures' Evidence and Gap Map in high income countries that include TIIPE. The Evidence and Gap Map is an interactive mapping resource that provides a visualisation of the type, quantity and quality of the available evidence regarding interventions designed to improve young people's employment and skills, and the relevant papers and reports that inform it.

This identified six recent evaluation reports published by the Youth Futures and 15 papers included in the Youth Futures Evidence and Gap Map. All published Youth Futures reports that include IPE were included. Papers from the Evidence and Gap Map were selected for inclusion in this review using the available filters and the default mode (OR within sections, AND across sections) as of the 30th October 2024.

- Country: High income countries
- Study confidence: Medium and high confidence primary study
- Study type: Process evaluation

Youth Futures reports

- Mackay, S., Gloster, R., Akhurst, E., Butler, N., McCoy, E., Timpson, H. & Quigg, Z. (2023). Evaluation of the St Giles Choices Programme: Final pilot evaluation report. Youth Futures, London.
- Beninger, K., Allingham, H., Mason, B. & Douch, W. (2024). The Diana Award Mentoring Programme Concept Test. Youth Futures, London.
- Beninger, K., Eastwood, L., Jouahri, S., Singh, S. & Karadia, A. (2023). An evaluation of the Stand Out youth employability programme. Youth Futures, London.
- Adams, L., O'Driscoll, C., Power, D., Cojocaru, M. & Kumari, M. (2023). CatZero Programme Concept Test. Youth Futures, London.
- Marren, C., Gray, H., Jacobsen, Y., Manetti, L., Robey-Turner, C. & Vaid, L. (2024). Evaluation of Youth Employment Service North East – final report. Youth Futures, London.
- Johnson, C., Downing, C., Power, D., Hassan, A., Wilkinson, A., Vaze, P., Marzano, L. & Aquije, H. (2023). Volunteer it Yourself evaluation. Youth Futures, London.

Evidence and Gap Map reports

- Brandenburg, U., Petrova, D., Bugarova, M., Kunc., M., Stiburek, S. & Tumova, P. (2016). The Erasmus impact study, regional analysis: a comparative analysis of the effects of Erasmus on the personality, skills and career of students of European regions and selected countries. European Commission.
- Coalter, F., Theeboom, M., Truyens, J., Soendgen, N., Gonzalez-Valles, E., Vukasinovic, N. and Vanden Berghe, S., (2020). Study on the contribution of sport to the employability of young people in the context of the Europe 2020 strategy: Final report. European Commission.
- Cutmore, M., Llewellyn, J. and Atkinson, I. (2020) Process evaluation of the essential life skills programme: Final evaluation report. Department for Education.
- Atkinson, I. et al. (2017) Youth Employment Initiative Process Evaluation: Assessment of Strategic Fit, design and implementation. Department for Work and Pensions.
- EKOS (2018) Evaluation of Youth Employment Initiative in Southwest Scotland. Scottish Government Managing Authority for European Social Funds (ESF).
- Jordan, L., McGinigal, S., Thomas, A. and Coleman, N. (2013). Early evaluation of the Youth Contract wage incentive scheme. Department for Work and Pensions.
- Ray, K., Crunden, O., Murphy, H. (2018). Liverpool City Region Youth Employment Gateway (YEG) Evaluation. Final Report. Liverpool City Region Employment and Skills Team.
- Peters, M. (2017) Helping unemployed creating their own work in the new economy: PES support to start ups. Publications Office of the European Union. European Commission.
- McCoshan, A. and Williams, J. (2002). Student apprenticeship evaluation. Department for Education and Skills.
- Chatzichristou, S., Ulicna, D., Murphy, I. & Curth, A. (2014). Dual education: A bridge over troubled waters? European Parliament's Committee on Culture and Education. European Commission.
- Nafilyan, V. and Speckesser, S. (2014). The Youth Contract provision for 16-and 17year-olds not in education, employment or training evaluation: Econometric estimates of programme impacts and net social benefits. Department for Education.
- Newton, B., Miller, L., Williams, J., Buzzeo, J., Hinks, R. and Gate, C. (2015). Process Evaluation of the Apprenticeship Trailblazers. Final Report, BIS Research Paper, (256). Department for Business, Innovation and Skills.
- Jeffrey, P. et al. (2020) Study for the evaluation of ESF support to youth employment. Publications Office of the European Union. European Commission.

- Roe, P., & Costello, M. (2014). Evaluation of shared apprenticeship pilots. Welsh Assembly Government Social Research. Welsh Government.
- Spielhofer, T., Mann, P. and Sims, D., 2003. Entry to Employment (E2E) participant study. Final Report, Slough, UK, National Foundation for Educational Research, NFER. Learning and Skills Development Agency (LSDA)

3. Definitions and concepts within TIIPE

This chapter focuses on RQ1: How is TIIPE defined and its function in different evaluation contexts described?

3.1. What is implementation and process evaluation (IPE) and how is it defined?

Several of the papers identified as part of this review offer definitions of IPE (e.g. Nielsen et al., 2022; Skivington et al., 2021; Cargo et al., 2018; Masterson-Algar et al., 2018; Humphrey et al., 2016a; Moore et al., 2015; Curran et al., 2012; Frye & Hemmer, 2012). Definitions are broadly aligned, and encompass consideration of the fidelity and quality of implementation (e.g. Skivington et al., 2021); the actions, functions and processes involved in implementation (e.g. Nielsen et al., 2022); the contextual and other influences on implementation (e.g. Curran et al., 2012; Moore et al., 2015), and explanation of how and why implementation succeeded or failed (e.g. Moore et al., 2015, Skivington et al., 2021, Humphreys et al., 2016a) and under what conditions (Humphrey et al., 2016a). In other words, IPE goes alongside and beyond effectiveness to consider a more holistic picture – including what might drive quality implementation.

Among the most comprehensive of these definitions is Humphrey et al.'s (2016a, p.3), who propose that IPE offers researchers "…*theoretical, methodological, and analytical tools that enable insights into the processes and mechanisms underpinning the impact (or lack thereof)*" of programmes. This wider focus includes a consideration of how a programme interacts with its implementation contexts, its contribution to wider systemic change, and how evidence can inform real-world decision making (Skivington, et al., 2021) and may also include what characteristics and policy supports are necessary for an effective intervention to scale (Humphrey, et al., 2016a).

Frye and Hemmer (2012, p. 298) similarly define IPE as the: "*systematic collection and analysis of information related to the design, implementation, and outcomes of a program, for the purpose of monitoring and improving the quality and effectiveness of a program*".

IPE has been described as a tool to enable researchers to "*look inside the black box*" and to tease apart how and why an intervention leads to change (Humphrey, et al.,

2016a). Moore et al. (2015) emphasise that IPE captures what is delivered in practice but also refers to the theory of the intervention in question.

3.2. What is theory in implementation and process evaluation (IPE)?

Some authors explicitly refer to theory within their definition of IPE and see the role of theory as central to IPE. However, across the reviewed texts, theory in the context of evaluation is conceptualised inconsistently, with authors referring either to programme theory, or implementation theory, or not making an explicit distinction between the two.

In its broadest sense, a theory is a set of analytical claims that structure the observation, understanding and explanation of the world (Seward, et al., 2021; Nilsen, 2015). According to Nilsen (2015, p.2), a 'good theory' provides clear explanations of how and why certain relationships cause certain events. Kislov et al. (2019) distinguish between different levels of theory in the social sciences, which include:

- **Grand theories** which are defined as overarching theoretical perspectives through which one can see and interpret the world. These are often formulated at a high level of abstraction. Examples include feminist theory and critical theory.
- **Mid-range theories** which are defined as the theories sitting between empirical observations and the programme theories based on them, and the more abstract 'view from on high' offered by grand theories. Mid-range theories can be further divided into '*lower-order theories*', e.g., theories which combine individual programme theories of similar interventions; '*higher-order theories*' which consolidate frameworks combining constructs from pre-existing mid-range theories, e.g., CFIR, (Consolidated Framework for Implementation Research; Damschroder et al., 2009) and TDF (Theoretical Domains Framework; Michie et al., 2005); and '*core implementation science theories*', e.g., Normalisation Process Theory (May & Finch, 2009) and the integrated Promoting Action on Research Implementation in Health Services (iPARIHS) framework (Harvey & Kitson et al., 2020).
- **Programme theories**, or 'small theories' which provide an explanation about how a specific policy, intervention or project is expected to function and achieve its objectives. Examples include programme theories of individual interventions or of large-scale knowledge translation initiatives, e.g. the National Institute for Health Research Collaborations for Leadership in Applied Health Research and Care (NIHR CLAHRCs) (Harvey et al., 2011).

In terms of the categorisation of theory, TIIPE most commonly operates at the level of programme theories and mid-range implementation theories, while often being informed by broader or 'grand' theories where these help to frame underlying

assumptions. At its core, TIIPE involves the articulation of assumptions, rationales and causal pathways about how programme components and/or their implementation are linked with outcomes (Wolfenden et al., 2021; Lafferty 2003; De Silva et al., 2014; Skivington et al., 2021; Ringhofer & Kohlweg, 2019). Pfadenhauer et al., (2017) explicitly reference the alignment between programme theory and implementation theory. They describe programme theory as an attempt to explain the causal mechanisms linking an intervention with its outcomes, and implementation theory as an attempt to explain the causal mechanisms of implementation (of an intervention).

Theory can also bring clarity to the complex systems in which implementation takes place (Damschroder, 2020) and it can provide explicit assumptions that can be tested, validated, or refined (Nilsen, 2015). Or, as De Silva et al. (2014) argue, understanding and documenting (e.g. via a theory of change) the theory underlying an intervention is key to understanding whether and how it works. A detailed theory of change can also help researchers to analyse where a programme theory is working and where it is not working during a phase of implementation (Ringhofer & Kohlweg, 2019).

Some authors note that evaluators have typically under-theorised and under-investigated how interventions interacted with their social contexts to bring about change (Fletcher, et al., 2016), and that the lack of explicit theories to explain how programmes are intended to enable change reduces the application of evaluation findings across contexts, making implementation "*an expensive version of trial and error*" (Eccles et al., 2005). Without bringing a theoretical lens to evaluation, evaluators are reduced to knowing about change in a single time and place, with learning difficult to generalise or apply elsewhere (Fletcher, et al., 2016).

3.3. How is TIIPE defined in the reviewed studies?

None of the included texts used the term 'theory-informed implementation and process evaluation'(TIIPE), nor did they offer an overarching definition that combines theory-based evaluation with implementation and process evaluation. Definitions were given for 'theory-based evaluations' – typically conceptualised as an approach that makes explicit the underlying programme theory and the conditions through which change is expected to occur. However, these definitions rarely extended to the integration of theory-based evaluation with IPE. As noted above, there were also few explicit references to the simultaneous use of programme theory and implementation theory when designing or conducting an IPE. This pattern is unsurprising as TIIPE is a term that is

used by Youth Futures but is not yet widely established in the broader evaluation literature.

3.4 How does incorporating theory add value to IPE?

There was broad consensus in the literature about the advantages of adopting a theory-informed perspective to evaluation. Sekhon et al. (2017) argue that the use of theory can add value across the development, evaluation, and implementation of interventions, and Nielsen et al. (2022) assert that adopting a theoretical approach to IPE ensures that the evaluation process is comprehensive and allows for comparisons between studies (Nielsen, et al., 2022). Nilsen (2015) also weigh up the advantages of applying a formal theory versus common sense (i.e. an informal theory), as summarised in this quote below:

"While the use of theory does not necessarily yield more effective implementation than using common sense, there are certain advantages to applying formal theory over common sense (i.e. informal theory). Theories are ***explicit*** *and* ***open to question and examination****; common sense usually consists of implicit assumptions, beliefs and ways of thinking and is therefore more difficult to challenge. If deductions from a theory are incorrect, the* ***theory can be adapted, extended or abandoned****. Theories are more* ***consistent with existing facts*** *than common sense, which typically means that a hypothesis based on an established theory is a more educated guess than one based on common sense. Furthermore, theories give individual facts a* ***meaningful context*** *and* ***contribute towards building an integrated body of knowledge****, whereas common sense is more likely to produce isolated facts". (Nilsen, 2015, p.9: emphasis added by current authors)*

A summary of other advantages of using theory in IPE discussed within the reviewed studies is given below:

- **Focusing evaluation activity**: Allowing evaluation to focus attention and resources on key aspects of an intervention (Lafferty & Mahoney, 2003).
- **Refinement of theory through iteration**: Enabling erroneous theoretical inferences to be adapted or discarded (Nilsen, 2015), facilitating the improvement of theories (Damschroder, Reardon, Widerquist, & Lowery, 2022; Skivington, et al., 2021).
- **Transparency and consensus**: Asking practitioners to make explicit and open assumptions that can be scrutinised, or encouraging practitioners to reach a

consensus with colleagues about what they are trying to do and why (Sekhon, Cartwright, & Francis, 2017; Nilsen, 2015; Lafferty & Mahoney, 2003).

- **Increased generalisability and transferability**: Generating richer learning that can be generalised and that enables better understanding of the contexts in which and conditions under which an intervention might work (Moullin et al., 2020; Skivington, et al., 2021).
- **Development of knowledge base**: Facilitating the accumulation and aggregation of evaluation results into a broader evidence base, and developing an integrated body of knowledge, by placing individual facts in a context of meaning (Nilsen, 2015; Lafferty & Mahoney, 2003).
- **Increased influence**: Increasing the degree to which evaluation results have an impact on policy and public opinion (Lafferty & Mahoney, 2003), by producing evidence that is useful to decision makers (Skivington, et al., 2021).

3.5. Theory in relation to programmes

It has become an expectation to include programme theory in evaluation designs (Wolfenden, et al., 2021). Across the reviewed texts, there was no single consensus definition of what a programme theory or theory of change should include, but there was broad alignment on several core elements: that programme theory should describe how an intervention is expected to generate its effects and under what conditions, identify key components and mechanisms of action, consider contextual influences, articulate pathways or causal chains through which change is expected to occur, and where possible – be informed by evidence, or established theoretical frameworks

Definitions of programme theory had varying levels of specificity. For instance, Seward et al. (2021) suggest that programme theory describes how a particular intervention is expected to yield its effects and under what conditions. Skivington et al. (2021) extend this definition to include an articulation of the key components of an intervention, the proposed mechanisms of action, and any contextual features that could impact the implementation of the intervention and its outcomes. Ringhofer & Kohlweg (2019) also specify that programme theories must be flexible to be useful.

In programme theory, the articulation of how a specific intervention leads to change is generally conceived as the programme theory of change. In other words, a theory of change breaks down or maps a programme from its inputs to outcomes by detailing a series of intermediate steps, such as activities and outputs, that, when articulated, form assumptions or hypotheses about how change occurs. These assumptions are also described as '*pathways of change*' or '*causal chains*' (Ringhofer & Kohlweg, 2019). In this

way, the assumptions of the programme are made explicit and therefore open to empirical verification (De Silva, et al., 2014).

Wolfenden et al. (2020) note that a theory of change can be developed using informal theory, that is, the application of tacit knowledge or common sense. However, it is usually recommended that programme theory is underpinned by wider evidence and/or a theory from a range of theoretical approaches, including formal theories or frameworks from behavioural or implementation science (Wolfenden et al., 2020; Skivington, et al., 2021).

Theories of change can also be articulated in narrative or visual forms. However, there is no single way to do this and there is a degree of variability in presentation and complexity (Humphrey, et al., 2016a). We discuss this further in Section 4.

Key aspects of programme theory are summarised in Table 3 and are described in more detail in the section that follows. Across the reviewed literature, there was no single consensus definition for all concepts in Table 3, but there was broad alignment on several core elements of programme theory and theory of change. The sources generally agreed that strong programme theory articulates the problem being addressed and the target population (De Silva et al., 2014), considers contextual influences (Skivington et al., 2021; Moore et al., 2015), identifies core components and mechanisms (Skivington et al., 2021; Moullin et al., 2020), and maps causal pathways and assumptions (Ringhofer & Kohlweg, 2019; Frye & Hemmer, 2012). Fletcher et al. (2016) and others are critical of approaches that focus on linear interactions and do not sufficiently address mechanisms of change. However, none of the texts stated that all concepts must be included for a theory of change to be considered valid; instead, they present a flexible set of commonly referenced elements.

Table 3: Programme theory key terms and definitions

CONCEPT	DEFINITION(S)
THEORY OF CHANGE	A Theory of Change (ToC) is an articulation and representation of a programme theory (De Silva et al., 2014) and is used to explain how and why a programme or intervention is expected to work, by highlighting the key factors representing programme success. It can be presented through a visual image which shows the expected cause-and-effect relationships within a specific context (De Silva et al., 2014).

CONCEPT	DEFINITION(S)
PROBLEM STATEMENT (RATIONALE)	The targeted challenges currently experienced by people, communities, organisations or systems. A ToC may also include a discussion of opportunities – these are the ways in which social conditions might be improved through the programme (De Silva et al., 2014).
TARGET POPULATION	The people, groups, organisations or systems the programme is designed to benefit, support or influence (De Silva et al., 2014).
CONTEXT	Elements external to the intervention such as social, organisational, cultural, political factors etc that may present a barrier or facilitator to the programme's mechanisms of action or expected outcomes (Skivington et al., 2021; Moore et al., 2015).
INTERVENTION	A policy, programme, project or action that seeks to achieve a desired outcome or impact through facilitating change or addressing root causes (Craig et al., 2018).
CORE COMPONENT	The essential, non-negotiable elements of a programme that are responsible for producing its intended effects. These components reflect the key activities, mechanisms, or structures that must be delivered with fidelity for the programme to function as theorised (Skivington et al., 2021).
MECHANISMS	Mechanisms are the underlying processes that explain how and why an intervention's components are expected to lead to the desired outcomes of the intervention (Skivington et al., 2021; Moullin et al., 2020).
CAUSAL PATHWAYS	The logical sequence of steps through which a programme's specific activities connect to outcomes and ultimately result in long-term impact (Ringhofer & Kohlweg, 2019).
OUTCOMES	The intended impacts, the markers of intervention success (Skivington et al., 2021; Moore et al., 2015).

CONCEPT	DEFINITION(S)
ASSUMPTIONS	Ideas, concepts and beliefs held by programme developers and stakeholders as to how an intervention will work in practice (Frye and Hemmer., 2012).

3.6. Key constructs used in programme theory

3.6.1. Context

Across the selected texts, context was consistently recognised as a central element of programme theory, with broad agreement that understanding contextual influences is critical to explaining how programmes interact with their environments and produce effects (Blamey and Mackenzie, 2007). Context refers to the various factors and conditions that surround an intervention or programme. It can include the external and internal environments, the characteristics of the target setting, stakeholders, and the broader sociopolitical or cultural landscape that can influence effectiveness (Damschroder et al., 2009). Although authors offered different classifications and emphases, they converged on the idea that context operates at multiple levels and may facilitate or constrain programme mechanisms and outcomes.

Authors recommend exploring contextual factors by adopting a multi-level or systems-level approach to context (Lengnick-Hall et al., 2023; Seward et al., 2021) and considering how they affect the achievement of the desired change (Cargo et al., 2018). This approach acknowledges and addresses programmes as a system that interacts with other systems within a broader context of delivery, and that can result in unanticipated outcomes (Seward et al., 2021). Skivington et al. (2021) recommend viewing programmes as events or interruptions in systems.

There are different accounts of the key dimensions of context that should be considered. For instance, Skivington et al. (2021) argue that, the physical, spatial, organisational, social, cultural, political, or economic dimensions of the context in which an intervention is implemented are crucial considerations. Kamphuis et al. (2022) and Pfadenhauer et al. (2012), both refer to context as constituted by geographical, epidemiological, socio-cultural, socio-economic, ethical, legal, and political characteristics that can influence an intervention, locally, nationally, and internationally.

Nielsen and Abildgaard (2013), propose that context can be understood through three broad interacting categories: omnibus context, discrete context, and organisational context – each of which shapes how an intervention is implemented and how outcomes emerge. Although framed within organisational research, these three categories align closely with what implementation science describes as implementation determinants – i.e., the contextual factors that influence whether implementation strategies activate their intended mechanisms.

- **The omnibus context** refers to the broad situational environment into which a programme is introduced. This includes the setting, timeframe, participants, purpose, and other background features that define the general conditions within which implementation occurs. It sets the stage for delivery by shaping the overall parameters, constraints, and expectations surrounding the intervention.
- **The discrete context** refers to specific events or situational factors – often unplanned or unpredictable – that can meaningfully influence implementation or outcomes. Examples include organisational restructuring, leadership turnover, introduction of parallel and/or conflicting programmes, policy shifts, or crises. These discrete factors can disrupt, accelerate, or redirect implementation processes and may introduce new mechanisms not anticipated in the original programme theory.
- **The organisational context** captures the local social and structural conditions that shape behaviour, decision-making and organisational functioning. This includes norms, relationships, communication patterns, readiness for change. It reflects the internal environment that influences how staff engage with the intervention and whether implementation strategies can be enacted as intended.

Pawson and Tilley (2004), categorise aspects of context as the "four I's", namely: 1) the *individual capacities* of programme participants and staff; 2) the *interpersonal relationships* created between them; 3) the *institutional setting* of the programme; and 4) the wider *infra-structural and welfare systems* within which the programme is embedded.

Context can act as both a facilitator and a barrier to impact, depending on how it aligns with or challenges the goals and strategies of the intervention (Greenhalgh et al., 2004).

Another distinction to be mindful of when considering context is its dynamism (Greenhalgh & Manzano, 2022). Some definitions of context conceptualise it as the backdrop in which implementation takes place at *one moment in time*. In contrast, other conceptualisations propose a more dynamic character, whereby context actively interacts with the intervention throughout its implementation (Greenhalgh & Manzano, 2022).

Approaches to identifying context

The precise contextual levels and features that will be appropriate to include in an individual IPE are highly dependent on the specific intervention being evaluated, thus, not every aspect of context and level will be relevant for every case (Craig et al., 2018). As such, evaluators should work with stakeholders to prioritise those aspects of context that are most consequential for the research questions and for understanding variability in outcomes (Skivington et al. 2021).

Identifying relevant contextual features

Craig et al. (2018) describe a range of contextual aspects that may influence intervention development, adaptation and evaluation, such as organisational structures, system-level conditions, and interpersonal dynamics. These features can act as moderators or mediators of implementation and outcome patterns and therefore require targeted examination.

Approaches for examining the role of context

To understand how contextual features shape implementation and outcomes, the literature highlights several methodological approaches, each offering a different analytical contribution:

1. **Realist approaches**: Realist evaluation emphasises the centrality of context for understanding how and why interventions work (or fail), highlighting variation across settings and subgroups (Pawson & Tilley, 1997; Craig et al., 2018). Realist logic supports identifying context–mechanism–outcome configurations that help explain for whom and under what conditions effects occur.
2. **Process evaluation:** Process evaluations can be used to examine the interaction between intervention activities and their contextual conditions. They can help identify which contextual influences enable or constrain delivery and how these factors shape participants' engagement and the unfolding of hypothesised mechanisms (Craig et al., 2018).
3. **Causal modelling approaches:** Causal models, such as mediation or moderation analyses, can be used to examine the pathways through which contextual variables influence outcomes. These approaches allow evaluators to test whether contextual factors mediate intervention effects or amplify, reduce or alter the direction of impact (Craig et al., 2018).

Measurement approaches for context

Wolfenden et al. (2020) note that context can be examined through either qualitative or quantitative measures. Qualitative approaches can reveal how actors experience contextual barriers or facilitators, while quantitative approaches allow assessment of the strength and distribution of contextual influences across settings. Used together, they contribute to a more comprehensive understanding of context within programme and implementation theory.

3.6.2. Programme mechanisms

Across the reviewed texts, mechanisms were consistently recognised as a central and necessary element of programme theory, and there was broad agreement that they should be investigated across evaluation phases (Czonek et al., 2022; Skivington et al., 2021; Nilsen & Abildgaard, 2013, Moore et al., 2015). However, despite this shared emphasis, the literature highlighted substantial variation in how mechanisms are defined and conceptualised, with authors noting ambiguity and the use of multiple, sometimes divergent definitions (Astbury & Leeuw, 2010; Lemire et al., 2020) dependent on the context and field of knowledge in which the term is being applied (Astbury and Leeuw, 2010). Lemire et al. (2020) draw out three distinct yet complementary definitions from Pawson and Tilley's work in realist evaluation, each focusing on different facets of how interventions lead to outcomes.

- **Mechanism as programme component:** Here, mechanisms reside in the programme and bring about change by introducing 'appropriate ideas and opportunities' for participants to respond to, e.g., providing advice (Pawson and Tilley, 1997). The programme introduces these mechanisms within specific settings and with specific people, essentially equipping participants with resources and knowledge that drive the intended outcomes.

- **Mechanism as participant reaction to programme components: Closely** aligned to the first definition, this perspective views mechanisms as emerging from the reasoning of participants and their engagement with programme activities. Here, 'mechanism' is conceived in two parts: the new resource being introduced by the intervention, together with, participants' attitudinal and behavioural responses to it (Pawson, 2004; Dalkin et al., 2015). Outcomes are shaped by how participants interpret and react to what the programme provides within a particular context.

- **Mechanism as explanatory account:** Lastly, Pawson and Tilley (2008) conceive of a mechanism as an explanatory account, a mini theory that articulates the make-up, behaviour and interrelationships between those processes that are responsible for outcomes, rather than as specific programme elements or participant reactions.

Astbury and Leeuw (2010) describe the three key characteristics of mechanisms as 1) usually being hidden, 2) being sensitive to context and 3) as generating outcomes. They draw on the work of Hedstrom and Swedberg (1998) to identify a possible typology of mechanisms operating at different levels of analysis:

- **Situational mechanisms (macro -> micro) –** these explain how features of a social or organisational context shape individuals' beliefs, motivations or opportunities, e.g., a new policy may alter an individual's motivation and how they respond to an intervention.
- **Action formation mechanisms (micro -> micro) –** these explain how individuals' choices and actions are shaped by specific combinations of beliefs, desires and opportunities, e.g., an individual's decision to adopt a new practice may be influenced by their perceived ability.
- **Transformational mechanisms (micro -> macro) –** these explain how multiple individuals' actions accumulate or interact to produce collective system-level outcomes, e.g., many staff members each making small behavioural adjustments collectively shifting organisational norms and practices and producing a larger structural change.

Approaches to identifying mechanisms

Some commentators suggest that qualitative methods are most suited to uncovering mechanisms since they can be hidden, complex and resistant to quantification (Renmans & Pleguezuelo, 2023). Interviews are especially common where the mechanisms in question relate to participants' behaviour and attitudes, enabling evaluators to understand participants' reasoning and responses to the intervention (Manzano, 2016). Observations are also used to capture interactions in context, although they often require supplementary qualitative data to reveal the underlying mechanisms fully.

Mediator analyses are also used to assess mechanisms (Craig et al., 2018). In this context, mediators are observable variables that are positioned between exposure to programme resources and an outcome, acting as measurable proxies for the underlying processes theorised to drive change. For example, if a programme seeks to improve outcomes by increasing participants' knowledge, "knowledge gain" can be measured as a mediator, even though the deeper cognitive processes that constitute the mechanism may remain unobservable. In this way, mediators provide a practical means of operationalising mechanisms within quantitative analyses.

Quantifying the contribution of these mediators helps evaluators understand not only whether an intervention works, but how much of its effect can be attributed to different pathways. This, in turn, strengthens understanding of how and why outcomes are produced – for instance, by showing whether certain mechanisms are more influential in some contexts than others, or whether hypothesised pathways are active at all. Such analyses can illuminate why effects differ across settings or subgroups. Mediation analysis, however, uses advanced statistical techniques and depends on accurately specifying the causal model and measuring all relevant variables. These demands are complex, and applications in public health intervention research remain limited (Craig et al., 2018).

3.6.3. Programme outcomes

There is broad agreement in the literature that outcomes represent the intended short-, medium-, and long-term changes resulting from a programme's activities. Although definitions varied in emphasis, the texts converged on the idea that outcomes may occur at individual, group, organisational, or system level and can include changes in knowledge, behaviour, health status, or wider social and policy environments (Frye & Hemmer, 2012; Skivington et al., 2021).

Programme outcomes (for beneficiaries of a programme) are distinguished from implementation outcomes (successful implementation of a programme), which we discuss in Section 5.

Adverse outcomes and consequences

Unforeseen and adverse outcomes may arise from a programme, making it essential for evaluators to anticipate, monitor and investigate any unintended consequences, particularly those that could cause harm (Seward et al., 2021). During the programme design phase, commentators advise that steps should be taken to consider potential uncertainties and ensure flexibility in addressing emerging challenges (Skivington et al., 2021). Evaluators must also recognise during the programme theory development phase the potential for interventions to exacerbate inequalities and disproportionately affect certain groups (Kamphuis et al., 2022). Engaging diverse stakeholders can enhance understanding of the mechanisms that may lead to harmful impacts on specific populations (Griffiths et al., 2022). Prompt action should be taken to address adverse consequences, which may involve redesigning the intervention, adapting implementation, or introducing new programme elements (Kamphuis et al., 2022).

Approaches to identifying and measuring programme outcomes

The literature emphasises the importance of using mixed methods - both qualitative and quantitative, to identify and measure outcomes effectively (Moore et al., 2015; Skivington et al., 2021). Quantitative methods are employed to measure effectiveness (Moore et al., 2015). Qualitative methods help explore, uncover, and explain outcomes, particularly those that are complex or context-specific, providing deeper insights into how outcomes are experienced and the underlying mechanisms that need to be present to create an outcome (Moore et al., 2015). Qualitative methods also allow emerging or unanticipated influencing factors to be identified (Lengnick-Hall et al., 2023; Humphrey et al., 2016) understanding how and why an intervention works in a real-world setting (Seward et al., 2021).

The literature also suggests it is important to consider both the research questions (Frye Hemmer, 2012) and the stage of evaluation when selecting methods to identify and measure outcomes. In formative evaluations, qualitative methods are particularly useful for identifying relevant outcomes and understanding how the programme works to produce them (Patton, 2015; Moore et al., 2015). Summative evaluations focus on measuring program impact, relying on specific quantitative indicators to assess effectiveness (Skivington et al., 2021; Craig et al., 2008). However, qualitative approaches provide valuable context to summative evaluations by explaining how and why certain outcomes occur, offering a more comprehensive understanding of programme effects (Greenhalgh et al., 2015).

3.7. Theory in relation to implementation

Across the reviewed texts, implementation theory was consistently recognised as an important lens for understanding how and why interventions are delivered successfully in real-world settings. In contrast to the literature on programme theory—where considerable variation in definitions and approaches was evident—there was more convergence in discussions of implementation theory. Authors broadly agreed that implementation theory should articulate the determinants, processes, and strategies that shape delivery, and guide the assessment of implementation outcomes such as fidelity and acceptability. There was broad alignment that implementation theory provides the explanatory grounding for understanding variation in delivery across contexts and therefore should be integrated into the design and interpretation of IPEs (Damschroder et al., 2009).

A range of theoretical approaches exists across different disciplines to guide in incorporating implementation theory within their evaluations (Nielsen et al., 2022; Ogden & Fixsen, 2014). To ensure a theory-informed approach offers as complete an

understanding and explanation as possible, evaluators should try to avoid narrowly focusing on theories from a single discipline, and draw on a range of approaches (Moore, et al., 2015; Nilsen, 2015). However, while multiple theories can enrich explanatory power, too many can risk conceptual overload or dilute the coherence of the resulting implementation theory.

Key aspects of implementation theory are summarised in Table 4. Further details about these aspects follows.

Table 4. Implementation theory key terms and definitions

CONCEPT	DEFINITION(S)
IMPLEMENTATION	The deliberate process by which an intervention is delivered within a setting (Pfadenhauer, et al., 2017; Moore, et al., 2015).
IMPLEMENTATION PROCESS	The social structures, resources, tactics, and mechanisms through which the implementation of an intervention is achieved (Moore, et al., 2015). The implementation process is characterised as active, iterative, and dynamic, and therefore non-linear (Pfadenhauer, et al., 2017).
IMPLEMENTATION DETERMINANTS	The factors that may help or hinder implementation (Damschroder et al., 2009).
IMPLEMENTATION STRATEGIES	The methods or techniques designed to support the implementation of an intervention (Wolfenden, et al., 2021; Moullin, et al., 2020).
IMPLEMENTATION MECHANISMS	The processes or events by which an implementation strateg(ies) operates to accomplish the desired implementation outcomes (Czonek, et al., 2022).
IMPLEMENTATION OUTCOMES	Proctor et al. (2011) broadly conceptualises implementation outcomes as the indicators of implementation success.

There has been a proliferation of theories, models and frameworks (TMFs) used to explain implementation, with a recent scoping review identifying 143 in public health and healthcare research alone (Wang et al., 2023). To bring clarity to this landscape, Nilsen (2015) provides a widely used categorisation that distinguishes between theories, models and frameworks:

- **Theories** provide broad explanations of phenomena and offer a conceptual understanding of how and why certain factors influence implementation. They tend to be general and grounded in fundamental principles.
- **Models** are more specific and prescriptive than theories, depicting relationships between variables and often with an emphasis on process, steps, or stages in implementation.
- **Frameworks** set out key constructs or dimensions of implementation and provide a structure for understanding and organising these elements. They are primarily descriptive and do not necessarily explain underlying mechanisms (Nilsen and Moore, 2024).

While theories can illuminate ***why*** certain factors affect implementation, models and frameworks are more practical, more often guiding ***how*** implementation should be considered and studied. Within this landscape, Nilsen (2015) identifies several types of TMFs relevant for IPEs, including:

1. TMFs for evaluating implementation

These frameworks specify aspects of implementation to examine and help to structure IPE designs. Examples include:

- **RE-AIM** (Aarons et al., 2011; Moullin et al., 2019), which recommends evaluating reach, effectiveness, adoption, implementation and maintenance.
- **The Implementation Outcomes Framework** (Proctor et al., 2011) which defines and conceptualises implementation outcomes such as acceptability, feasibility and fidelity. These outcomes are often used as anchors for TIIPE.

2. TMFs to understand and/or explain what influences implementation outcomes

Nilsen further distinguishes between:

- **Classic theories:** Theories drawn from psychology, sociology and organisational sciences. These explain how change occurs and what influences behaviour. They may address issues such as organisational behaviour (including e.g. leadership, organisational learning or organisational culture), other collective behaviour (e.g. social networks and social capital), or individual behaviour (e.g. social cognitive theory).

- **Implementation theories:** Developed or adapted specifically to explain implementation phenomena. Examples include:
    - **Normalization Process Theory (NPT)** (May et al., 2009) which focuses on the process by which new practices become normalized and integrated into everyday work routines
    - **COM-B** (Michie et al., 2011) which posits that the behaviour change required for effective implementation requires capability, opportunity and motivation.
    - **Organisational Readiness Theory** (Weiner, 2009) – emphasising collective commitment and confidence in implementing change.
- **Determinant Frameworks:** These identify multilevel factors (determinants) that influence implementation outcomes (Czonek, et al., 2022; Woodward, et al., 2021). They overlap substantially with contextual considerations in programme theory. Determinant frameworks help evaluators identify relevant contextual conditions, barriers and facilitators, and can guide the development of context measures (Damschroder, et al., 2022). Examples include:
    - **Consolidated Framework for Implementation Research** (CFIR) (Damschroder et al., 2009; 2022b)
    - **Integrated Promoting Action on Research Implementation in Health Services (i-PARIHS)** framework (Harvey & Kitson et al., 2020)
    - **Theoretical Domains Framework (TDF)** (Michie et al., 2005)
    - **Exploration, Preparation, Implementation, Sustainment (EPIS)** framework (Aarons et al., 2011; Moullin et al., 2019)

Nilsen (2024c) provides examples of the application of EPIS, CFIR, i-PARIHS, COM-B, NPT and RE-AIM in practice, and illustrates how the use of these TMFs can enhance explanatory depth in evaluations and supports more structured, transparent, and theory-driven IPEs.

3.8. Key constructs used in implementation theory

3.8.1. Implementation outcomes

Implementation is a process, and its quality can be measured using several indicators, i.e., a construct called implementation outcomes. These outcomes should be defined, and be the focus, in IPE (Proctor, et al., 2023).

There are several implementation outcomes defined in the literature reviewed, which include the following outcomes highlighted in Proctor et. al's (2011) seminal paper, Damschroder et al. (2022) and Moore et al. (2015):

- Acceptability
- Adoption
- Appropriateness
- Feasibility
- Fidelity
- Implementation cost
- Penetration
- Reach
- Sustainability

The following section provides definitions for the implementation outcomes found in this literature review, as well as guidance about aspects of the constructs to be researched and/or measures for doing so.

Acceptability

Acceptability refers to the degree to which an intervention is perceived to be agreeable, palatable or satisfactory (Proctor et al., 2023). It is a dynamic construct that may shift over time, so the literature recommends assessing it at different timepoints during delivery (Sekhon, Cartwright, & Francis, 2017). It can be measured at the level of individual providers or recipients, using observed behaviours or self-report methods such as surveys, interviews, and administrative data (Sekhon et al., 2017; Proctor et al., 2011; Lafferty & Mahoney, 2003).

Adoption

Adoption, also known as uptake, is defined as the decision by individuals, organisations, or systems to initiate the use of a new intervention, programme, or practice (Proctor et al., 2011). This decision is influenced by various factors, such as perceived value, fit with existing practices, resources, and the readiness of the adopter (Damschroder et al., 2009; Rogers, 2003). It is an essential stage in the implementation process because without adoption, no further steps in implementation can occur (Rogers, 2003).

It can be measured through survey data or administrative data, tracking number of individuals/teams/organisations that have agreed to adopt the intervention (Holtrop, et al., 2021). Interviews or focus groups can also help understand the underlying reasons for adoption decisions, including attitudes, perceptions and contextual factors (Proctor et al., 2011).

Appropriateness

Appropriateness relates to the perceived fit, relevance or compatibility of an intervention within a specific context, including its alignment with the needs, priorities, values, and goals of the individuals, organisations, or communities involved (Proctor et al., 2011). It reflects the degree to which the intervention is considered suitable or justifiable in its chosen setting. It is distinct from acceptability as an intervention can be acceptable (i.e., people like it and see the value behind it), but inappropriate (e.g., because it is too complex to deliver within a low-resourced setting).

Appropriateness is often assessed through a combination of qualitative and quantitative data collection tools. Structured surveys or questionnaires can assess the degree to which stakeholders perceive an intervention as appropriate for their specific context (Proctor et al., 2011). Interviews and focus groups can provide deeper insights into why an intervention is, or is not, considered appropriate, exploring the underlying factors that shape stakeholders' perceptions (Damschroder et al., 2009).

Feasibility

Feasibility (or 'intervention feasibility' to distinguish it from the feasibility of an impact evaluation) is conceptualised as the extent to which an intervention can be successfully used or deployed in a setting (Proctor et al., 2023). It is concerned with whether the implementation of an intervention is practical and achievable within real-world conditions. Feasibility is a critical precursor to understanding whether an intervention can be effectively implemented on a larger scale (Proctor et al., 2011). By addressing feasibility concerns, organisations can increase the likelihood of successful intervention delivery and sustainability.

Key aspects of feasibility include resource availability, staffing, training, time, and support systems that enable the intervention to be put into practice without overwhelming the organisation or its staff (Damschroder et al., 2009). Surveys can be used to assess these logistical aspects of feasibility, and these can then be used to quantify barriers to implementation (Proctor et al., 2011). Qualitative assessments, such as interviews, focus groups, or site visits, can also provide detailed insights into the practical barriers and challenges to implementing an intervention (Damschroder et al., 2009).

Fidelity

Fidelity is a core implementation outcome, referring to the degree to which an intervention is implemented as it was designed or intended (Carroll et al., 2007; Proctor, et al., 2023). Fidelity is important because even the most effective interventions may fail to produce the desired outcomes if not delivered with sufficient quality and consistency (Carroll et al., 2007).

Fidelity includes several dimensions that help assess how closely an intervention matches its intended design. Carroll et al. (2007) set out five components of fidelity, which are referenced in other papers throughout the review:

- **Adherence:** the degree to which the intervention's activities or components are carried out as planned. This includes following protocols, maintaining key processes, and ensuring that the intervention is delivered to the intended target population.
- **Dosage**: The amount or intensity of the intervention delivered. This includes factors such as frequency, duration, and consistency in the delivery of intervention components.
- **Quality of delivery**: The way the intervention is delivered, including the competence, training, and skill level of those implementing the intervention. It also includes how well the intervention fits with local practices and the extent to which it is delivered with high standards.
- **Participant responsiveness**: The degree to which participants (e.g., individuals, patients, or community members) are involved and engaged in the intervention. Engagement is crucial for ensuring that the intervention's processes and intended outcomes are achieved.
- **Programme differentiation**: This aspect concerns identifying the unique aspects of an intervention which contribute to its intended outcomes, and without which, the intervention would not be successful, or in other words, identifying a programme's essential components.

Fidelity can be assessed through direct observation (e.g. in site visits, video recordings, or structured observation) or self-report (e.g. completing logs, surveys or interviews with staff or participants) to determine whether the intervention is being implemented as intended (Carroll et al., 2007; Proctor et al., 2011). Structured fidelity checklists or scales can be used to measure how closely the intervention is being delivered to the prescribed model. These tools help standardize assessments and allow for comparisons across sites or time points (Carroll et al., 2007).

There is a recognised tension between fidelity and adaptation (Skivington et al., 2021). Adaptation refers to modifications made to an intervention to better fit the local context and is an important part of implementation. There is a careful balance needed

between these two outcomes – adaptation may be necessary for local relevance, but too much adaptation can compromise fidelity and the intervention's effectiveness (Carroll et al., 2007).

Some authors advocate a focus on function over form in assessing fidelity, in other words, considering whether the key purpose of an activity was achieved, regardless of its superficial form, (e.g., whether a meeting takes place in-person or remotely (Skivington et al., 2021). This approach supports more flexible intervention designs that are responsive to participants' needs and preferences, allowing for adaptation without compromising the core aims of the intervention (Skivington et al., 2021).

Implementation cost

Implementation cost refers to whether the specific financial burden of an intervention effort makes it viable (Proctor et al., 2023). It is affected by the complexity of the intervention, the complexity of the implementation strategy, and the delivery setting (Proctor et al., 2023). It is measured at the level of the provider or delivering institution, usually using administrative data. (Proctor et al., 2011).

Penetration

Penetration refers to the integration of an intervention within a setting (Proctor et al., 2023). It captures how well it has been incorporated into routine practice or systems. Penetration is often considered an indicator of the scale-up and longevity of an intervention (Damschroder et al., 2009; Proctor et al., 2011).

One of the primary ways to measure penetration is through coverage, for example measuring the proportion of providers delivering a new practice within a setting or system (Proctor et al., 2011). Provider organisations' continued engagement with an intervention over time is an indicator of deeper integration and penetration (Proctor et al., 2011). Penetration can also be assessed through surveys or interviews with key stakeholders to understand how well the intervention has been adopted and integrated within the broader organisational or community context (Damschroder et al., 2009).

Reach

Reach refers to the extent to which an intervention or programme is delivered to and engages with all the intended target population (Proctor et al., 2011), or the proportion of individuals intended to be exposed to the intervention who do participate and their characteristics (Holtrop et al., 2021; Proctor et al., 2011). It is a particularly important consideration in relation to equity (considering which parts of an intended population are and are not reached) and very relevant to considerations of scalability.

Surveys, interviews, registration data or usage data can provide quantitative data on reach (Holtrop et al., 2021). Qualitative methods, such as interviews or focus groups, can help identify factors that hinder reach, such as barriers to access, lack of awareness, or logistical challenges (Damschroder et al., 2009).

Sustainability

Sustainability refers to the ability of a setting into which an intervention is embedded to continue to deliver and maintain its effectiveness after initial implementation efforts have ended (Proctor et al., 2023). In the RE-AIM framework (Holtrop et al., 2021) it is referred to as maintenance, considering whether the programme is still ongoing in settings, being delivered with fidelity, and whether intended impacts continue to be achieved. The length of time an intervention continues to be delivered is a basic measure of sustainability. Tracking whether the intervention remains in place over months or years provides insight into its ongoing viability (Scheirer, 2005).

This may be ascertained through interviews, surveys, or organisational records, complemented by assessment of the level of institutional or staff support for continued implementation. High levels of staff engagement and organisational commitment are indicators of sustainability (Aarons et al., 2011).

3.8.2. Implementation determinants

Implementation determinants are the factors that influence or shape the process and outcomes of implementing an intervention, or the factors that may help or hinder implementation. They can be modifiable (e.g., leadership support) or non-modifiable (e.g., community context) and operate at different levels, including the individual, organisational, and system levels (Proctor et al., 2009; Damschroder et al., 2009 and 2022b). Understanding implementation determinants helps practitioners and researchers identify barriers and facilitators to successful implementation.

Overlap between implementation determinants and context in programme theory

There is an overlap between context in programme theory and implementation determinants. The connection lies in their shared focus on external and internal factors that shape how interventions are implemented and their outcomes (Pfadenhauer et al., 2017). Both concepts recognise that the success of an intervention is influenced by contextual elements such as organisational culture, leadership, available resources, policy landscape (Pawson & Tilley, 1997; Damschroder et al., 2009).

Identifying implementation determinants

Several frameworks exist to help categorise and understand implementation determinants, offering a systematic approach to identifying and addressing factors that influence implementation.

Developed and revised by Damschroder et al. (2009, 2022b), the Consolidated Framework for Implementation Research (CFIR) categorises implementation determinants within five major domains:

- **Intervention characteristics:** or features of the intervention itself (e.g., complexity, adaptability)
- **Outer setting:** including external factors such as policy, resources and local conditions
- **Inner setting:** or organisational context, including infrastructure, relationships, culture, tension for change
- **Individuals:** including leaders, those implementing the intervention and 'recipients' and considering need, capability, opportunity and motivation
- **Implementation process** itself, including assessing needs, planning engaging stakeholders, and evaluating progress.

Another commonly used determinant framework is the Exploration, Preparation, Implementation, Sustainment framework, widely referred to as the EPIS (Aarons et al., 2011). EPIS defines four phases in the implementation process and documents factors that influence implementation at each stage. These factors are organised as those related to the outer system (i.e. the wider context outside the delivery organisation), the inner organisational context, and 'bridging factors' that link the outer system with the inner context, e.g., intermediary roles than span across system levels. EPIS also includes 'innovation factors' which relate to the characteristics of the intervention itself, e.g., how complex an intervention is to deliver.

As a further example, the Theoretical Domains Framework (TDF) identifies the behavioural determinants that influence professional practice. It includes domains such as knowledge, skills, beliefs, motivation, and environment (Michie et al., 2005).

Several determinant frameworks, including CFIR, EPIS and TDF have been developed with sets of qualitative and quantitative measures or questions as a recommended starting point to operationalising them in evaluations. In the case of CFIR, Damschroder et al (2022b) suggest that data on determinants is best collected directly from stakeholders within the implementation setting, such as key decision makers, those delivering the intervention, and its recipients.

Using determinants frameworks in evaluation can help to identify barriers and facilitators, using shared language and building a body of evidence. However, it is also recognised that using determinant frameworks can risk introducing bias into the research process (Nilsen, 2015), because by investigating pre-determined factors, evaluators may overlook others that also influence implementation outcomes. Implementation determinants are often context-specific and can vary across different settings, making it challenging to generalise findings from one context to another (Damschroder et al., 2009). Determinants such as organisational culture or staff beliefs are also difficult to measure directly. Researchers often rely on proxies or self-reports, which can introduce bias (Aarons et al., 2011). Determinants also often interact with each other, making it difficult to isolate their individual effects (Glasgow et al., 2019).

3.8.3. Implementation strategies

Implementation strategies are planned activities that aim to support the adoption, fidelity, and sustainability of an intervention (Powell et al., 2015). These strategies may include approaches such as training and education, technical support, and professional coaching, or financial incentives (Powell et al., 2015). The goal of implementation strategies is to improve implementation by addressing barriers or catalysing enabling factors (Powell et al., 2015).

The focus of TIIPE in terms of exploring implementation strategies, is to assess what strategies were used and whether they helped to facilitate the implementation process, including at different implementation stages (Moullin et al., 2020).

3.9. Integrating programme theory and implementation theory

In TIIPEs, testing theory involves testing the integrity of the programme theory itself and testing the implementation of the programme in context (Fletcher et al., 2016; Cargo et al., 2018). Many of the reviewed texts integrate discussion of programme theory and implementation, either explicitly or more often implicitly, or refer to the need to bring them together.

Several authors emphasise that programme theory should incorporate implementation considerations from the outset, so that assumptions about how the intervention works are explicitly linked to how it will be delivered in real-world conditions. Skivington et al. (2021) and Moullin et al. (2020) propose that early consideration of implementation benefits programme design and programme theory by ensuring causal pathways are plausible given the system in which the intervention will operate.

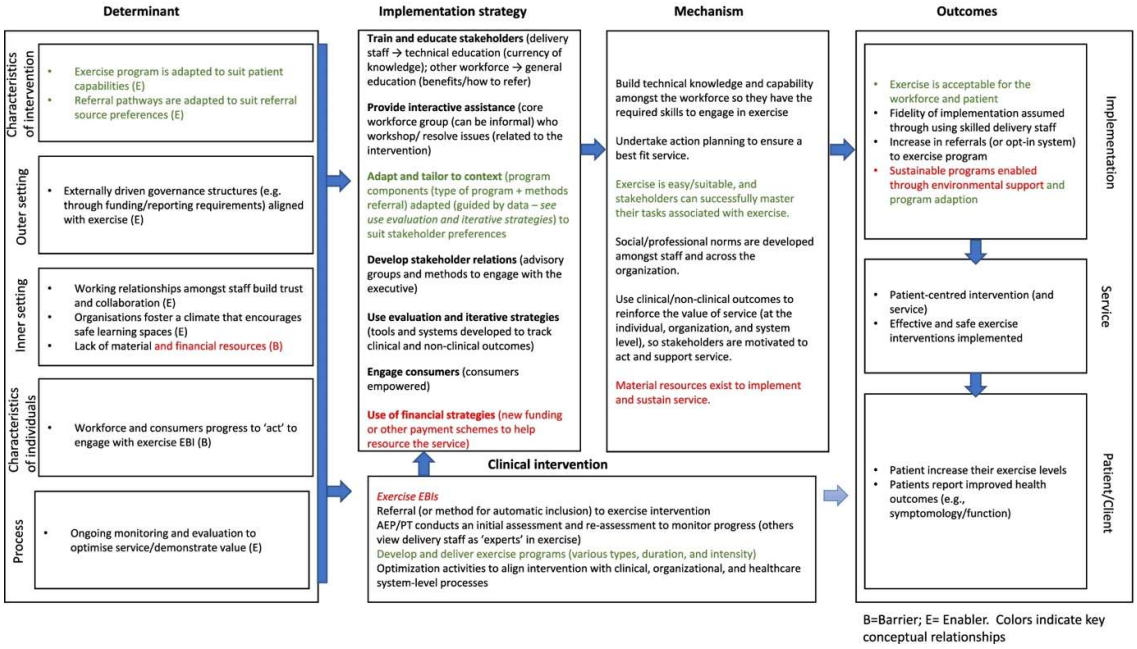
Cargo and colleagues (2018) provide one of the clearest illustrations of this integration. Their argument that evaluators must "*assimilate evidence*" to distinguish between programme theory success/failure and implementation success/failure by jointly interrogating programme theory and implementation theory. Testing these together involves structuring data collection to answer not only whether the outcomes predicted by the programme theory occur, but also whether implementation strategies were enacted as intended, how contextual factors influenced delivery, and how both intended and unintended outcomes arose through feedback loops, interactions and double-loop learning (Nielsen et al., 2023; Frye 2012).

Several authors highlight methodological tools that support this kind of integrative testing. De Silva et al. (2014) note that ToCs naturally integrate programme theory and implementation processes, as they allow evaluators to link activities, mechanisms, and contextual influences with both implementation and outcome data. Renmans & Pleguezuelo (2023) advocate the use of analytic methods such as structural equation modelling or systems modelling to test interconnected causal and implementation pathways within complex interventions.

From the implementation science literature, Smith et al. (2020) and Czosnek et al. (2022) propose the Implementation Research Logic Model (IRLM) as a complementary tool that explicitly maps implementation determinants, strategies, mechanisms and implementation outcomes, clarifying how these interact with (rather than replace) the intervention's programme theory (See example below, Figure 2). By distinguishing the intervention from its implementation system, the IRLM strengthens evaluators' ability to test how implementation theory supports, constrains or modifies the causal pathways articulated in the programme theory.

Taken together, the reviewed literature demonstrates that neither programme theory nor implementation theory is sufficient on its own for TIIPE. It is the *integration* of the two - linking hypothesised mechanisms of change with delivery realities, contextual conditions and implementation strategies that enables evaluators to robustly explain how and why interventions succeed or fail across different settings.

Figure 2. Example of an IRLM from Czosnek et al., (2022)



4. How to develop, test and refine theory in IPE

4.1. Overview

This chapter focuses on RQ2: How should evaluators approach the development, testing and refinement of programme theory and implementation theory?

It begins by outlining key principles for developing theory, including grounding a theory in existing evidence and drawing on the expertise of diverse stakeholders. It then examines how programme theories are commonly articulated, focusing on ToC as the predominant approach, while also considering alternatives such as logic models.

The chapter then turns to how both programme theory and implementation theory can be empirically tested and refined, outlining practical considerations for identifying uncertainties, prioritising what to examine, and assessing how programmes and implementation strategies interact with context. Finally, it draws together a set of cross-cutting principles - including the use of multiple constructs, multi-level contextual analysis, mixed methods and shared standards of rigour that support robust theory testing and ongoing refinement within TIIPE.

4.2. Developing programme theory

Across the reviewed texts, there was broad agreement on key principles such as grounding theory in evidence and involving a wide range of stakeholders in developing programme theory. While specific methods varied, the literature converged on the importance of drawing on existing research, practitioner knowledge and stakeholder experience to strengthen the plausibility and usefulness of the resulting programme theory.

Although the literature reviewed predominantly focuses on these principles in relation to programme theory, similar considerations will apply when developing an implementation theory, including drawing on existing theories about implementation (e.g., implementation frameworks or wider theory about organisational change or individual behavioural change).

4.2.1. Grounding the theory in evidence

In terms of 'what' should be involved when building programme theory, in the first instance evaluators are encouraged to use existing evidence and research literature to help them develop programme theories that are more likely to be borne out (Romão et al., 2023; De Silva et al., 2014; Craig et al., 2018; Skivington et al., 2021; Moore et al., 2015). This includes evidence from similar interventions alongside field-specific literature and theory. The theory-building process is described as explicitly cumulative and iterative with learning slowly accreted over time (Blamey and Mackenzie, 2007).

4.2.2. Building on the expertise of a wide range of stakeholders

There was general agreement on the importance of involving a wide range of stakeholders, including those using, delivering, evaluating, and funding the intervention (Romão et al., 2023; De Silva et al., 2014; Craig et al., 2018; Skivington et al., 2021; Moore et al., 2015). Building on a wide range of experience and expertise during theory development enhances its quality and relevance, helping to establish perceptions of impact through consensus building. The involvement of key stakeholders can support community engagement and ownership of a programme and help to build evaluation capacity (Blamey and Mackenzie, 2007). Involving multiple stakeholders can also help in the prioritisation of key issues e.g. areas of uncertainty for the evaluation to explore (Moore et al., 2015).

Considerations should be given to the:

- **Level or extent of involvement** (from extensive community-based participatory approaches to briefer approaches (Gustafson et al., 2023)
- **Methods and engagement activities used for sharing knowledge and expertise**, which could include participatory action research, participatory theory of change development and human-centred design (Seward et al., 2021) or using workshops, interviews and focus groups

4.3. Using theories of change to support TIIPE

Once the core elements of a programme theory have been developed, evaluators require a clear and systematic way of articulating these components. A ToC is the most common approach for doing so. ToCs provide a structured and often visual representation of how and why a programme is expected to work, outlining its rationale, activities, mechanisms, assumptions, contextual influences and outcomes. In doing so, they act as a practical bridge between the development of programme theory and its

subsequent testing and refinement within TIIPE, including the development of linked implementation theories. An example of a ToC, which includes an explicit representation of how and why an intervention works through causal mechanisms and a clear exploration of its rationale and assumptions, is presented in Figure 3.

While ToCs *can* be developed at any time up to the point of summative evaluations, it is best practice to develop one at programme design or inception. ToCs can then be refined iteratively.

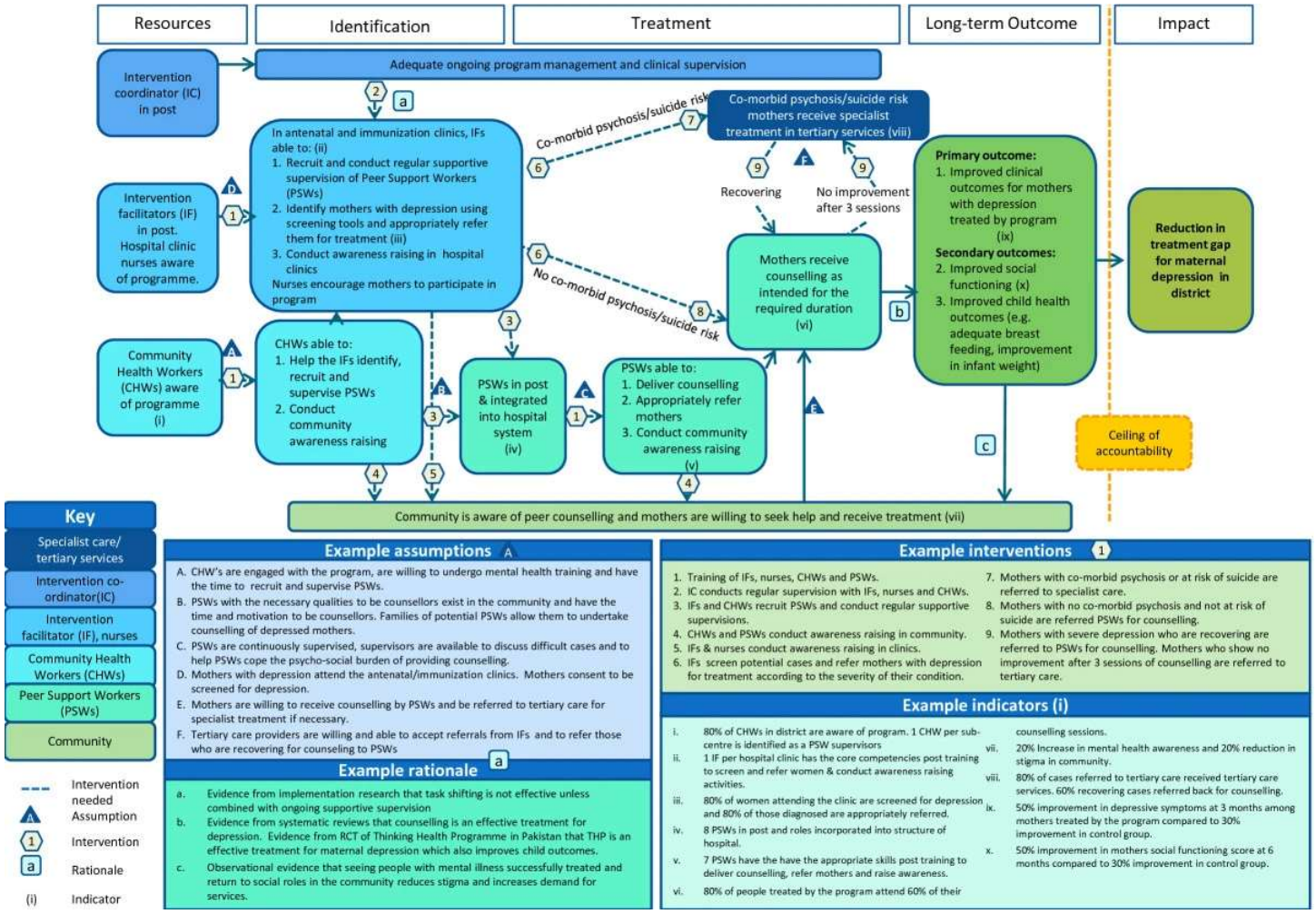
There is variation within the reviewed research literature on the order in which the components of a ToC should be developed. Romao et al., (2023) advocate working through defining the problem, expected outcomes, interventions and change mechanisms, before modelling, validating and revising the theory of change.

An alternative approach to visualising programme theory is logic models which capture an intervention's programme theory in a logical, stepwise fashion (Woodland and Mazur, 2024; Pfadenhauer et al., 2017; Kamphuis et al., 2022; De Silva et al., 2014; Czosnek et al., 2022) usually in the form of a flow diagram characterised by boxes and arrows. For example, a logic model might show that the input in Box A is intended to lead to the output in Box B, which in turn leads to the outcome in Box C.

Logic models have been criticised as being a limited and linear conceptualisation of causality and over-simplifying linkages between the activities of a programme and causal mechanisms and outcomes (Frye and Hemmer, 2012; De Silva et al., 2014; Fletcher et al., 2016). Some authors see ToCs as more powerful than logic models for articulating and visualising programme theories, on the basis that they can incorporate non-linear causal pathways (e.g. through feedback loops) and wider contextual influences and incorporate programme mechanisms, providing a more nuanced view of how interventions work (Frye & Hemmer, 2012; De Silva et al., 2014; Fletcher et al., 2016).

However, contextual factors and programme mechanisms are sometimes included in logic models (Woodland and Mazur, 2024; Czosnek et al., 2022), and the distinction between ToCs and LMs is not consistently described in the literature. Skivington et. al (2021) contend that what is of concern is not necessarily whether a programme theory is labelled a ToC or a logic model, but whether it includes the core components of programme theory as defined above (i.e. problem definition, intervention, implementation activities, outcomes, mechanisms).

Figure 3. An example Theory of Change for a peer counselling intervention for maternal depression in Goa, India - from De Silva et al. (2014



4.4. Testing and refining programme theory

Evaluators can use ToCs to guide the empirical testing and refinement of programme theory. Across the reviewed literature, testing theory is presented as a core function of IPE, although detailed practical guidance is limited. A shared message, however, is that theory should shape data collection, analysis and interpretation, and that theory refinement should be iterative and responsive to emerging evidence.

This section outlines how evaluators can test and refine programme and implementation theories and is followed by a set of cross-cutting principles to support this work.

4.4.1. Identifying uncertainties and prioritising what to test

Once developed and mapped, the programme theory guides the questions to be asked within a TIIPE. This involves testing whether the intended components of the programme were delivered, whether the intended outcomes are achieved, and whether there is evidence to support the hypothesised mechanisms. Evaluators are encouraged to focus on the most critical questions needed to understand the interplay between a programme, its implementation, and its context, rather than attempting to address all potential uncertainties - an approach that risks diluting resources (Skivington et al., 2021). Determining these core questions requires careful consideration of where uncertainties in the programme theory lie (DeSilva et al., 2014; Moore et al., 2015; Skivington et al., 2021, Craig et al., 2018, Durlak & DuPre, 2008; Frye & Hemmer, 2012). Consulting with wider policy and practice stakeholders can help identify which uncertainties are most essential to investigate. Whilst establishing these core questions early on sharpens the evaluation's focus, retaining some flexibility to address unexpected issues is important for capturing new insights that emerge during the evaluation (Moore et al., 2015; Skivington et al., 2021).

4.4.2. Using realist evaluation and contribution analysis to test programme theory

Several authors recommend enhancing theory quality by using methodologies that support deeper examination of causal processes and provide a more nuanced understanding of complex social change processes than evaluation approaches based solely on ToC or logic model approaches (Breuer et al., 2016).

For example, several authors emphasise the role that realist evaluation can play in the testing and refining of causal links and increasing explanatory power (Blamey and Mackenzie, 2007; Bruer et al., 2016; Fletcher et al., 2016; Renmans and Pleguezuelo, 2023; Rolfe, 2019; Skivington et al., 2021). Realist Evaluation can be used to identify and

test hypotheses about *how* and *why* the intervention works, *for whom*, and *in what circumstances.*

Similarly, Contribution Analysis is an approach that can be used for appraising causality, through building a credible and robust narrative about how a programme's activities lead to the desired outcomes (Mayne, 2011; 2019). Contribution Analysis tests the causal links between activities and outcomes, and the assumptions underlying them, assessing the strength of evidence supporting the contribution claims using multiple different sources of evidence.

4.4.3. When and how to refine theory

Refining programme theory (and implementation theory) is not a one-off step, but a continuous process, responding to what the evaluation uncovers. Most commentators contend that decisions guiding data collection, the approach to analysis and the interpretation of data should all be theory-led (Romão et al., 2023; De Silva et al., 2014; Craig et al., 2018; Skivington et al., 2021; Moore et al., 2015). However, little guidance is provided on precisely how to go about refining theory in the light of evidence, beyond that this should be based on the priority concerns for the study and input from wider stakeholders (Skivington et al., 2021).

4.5. Testing and refining implementation theory

In parallel with testing programme theory, evaluators are encouraged to empirically examine the assumptions embedded within an implementation theory - particularly hypotheses about mechanisms, determinants, and strategies. Authors emphasise that implementation theory should, like programme theory, be treated as provisional and subject to empirical examination, with evaluators using theory to develop claims or hypotheses about how implementation unfolds and then testing these claims against data (Seward et al., 2021). This includes assessing whether implementation strategies activate the anticipated mechanisms and whether contextual barriers or facilitators align with theoretical predictions (Damschroder et al., 2009; Pfadenhauer et al., 2017). The literature also highlights that implementation processes are shaped by dynamic interactions between context, mechanisms, and strategies, meaning that evaluators need to examine how these factors interplay in practice (Greenhalgh & Manzano, 2022). Treating implementation theory as something to be empirically tested in this way strengthens both its explanatory power and its usefulness for guiding adaptation and scale-up.

4.5.1. Methodological approaches: SMART trials for testing and adaptation

One emerging approach for testing and refining implementation theory is the use of SMART trials (Sequential Multiple Assignment Randomized Trials). SMART designs allow evaluators to determine which implementation strategies (or combination of strategies) work best for different groups or contexts and how adaptations during implementation affect outcomes (Collins et al., 2007). The methodology involves randomly assigning participants (or settings) to various implementation strategies and then randomising again to tailor implementation efforts based on real-time feedback. Unlike traditional randomised controlled trials (RCTs), which typically involve a fixed intervention, SMART trials provide flexibility to respond to emergent implementation outcomes or new data, making them particularly useful for dynamic real-world settings where conditions evolve (Murphy, 2005) and enabling iterative refinement of implementation theory.

4.6. Cross-cutting principles for testing programme and implementation theories in IPE

The preceding sections focus on how programme and implementation theories can be articulated and directly tested. The literature also points to several cross-cutting principles that support rigorous theory testing and refinement. These principles apply to both programme theory and implementation theory.

Inclusion of multiple theoretical constructs

The literature emphasises that robust IPE requires attention to multiple theoretical constructs across both programme theory and implementation theory. Evaluators are encouraged to examine not only implementation outcomes, determinants, and strategies (e.g., Humphrey, et al., 2016a), but also the key mechanisms and assumptions embedded within the programme theory that explain how change is expected to occur. Assessing a range of constructs helps identify which aspects of programme design and delivery are most critical for success, while also recognising that not every construct will be feasible or relevant to measure in every evaluation.

Guidance on selecting constructs stresses the importance of drawing on expert consensus, previous empirical work and studies, and relevant policy or change theories (Damschroder et al. (2022). From an implementation perspective, core constructs often

include reach, dosage, and fidelity - alongside co-intervention, countermeasures and contextual factors (Cargo et al., 2018). From a programme theory perspective, evaluators should prioritise the mechanisms and assumptions most central to the hypothesised causal pathway, selecting measures that help determine whether these mechanisms are activated as intended.

Mapping and operationalising multi-level contexts

Because both programme theory and implementation theory assume that context shapes mechanisms and outcomes, evaluators are encouraged to examine contexts at multiple levels when testing theoretical propositions e.g. individual, organisational and system levels and involving all organisations and participants involved in an intervention's delivery. Addressing context as a single-level issue of implementation is only appropriate in extremely rare cases (Lengnick-Hall et al., 2023).

Mixed methods as standard practice

Testing theory requires methods capable of capturing different types of evidence about how programmes work and how they are implemented in practice, and mixed methods combining qualitative and quantitative data are recommended as standard practice (Seward et al., 2021; Humphrey et al., 2016). However, authors also acknowledge that budgets and other constraints may hinder this and that evaluators should strike the "*optimal balance between rigour and feasibility*" (Humphrey et al., 2016).

Qualitative methods are considered more appropriate to capture aspects such as the experiences of key stakeholders, how change occurs and emergent modifications to the implementation process (Cargo et al., 2018; Humphrey et al., 2016; Moore et al., 2015). As well as generating new and potentially surprising insights, qualitative methods are an important tool for validating or modifying the intervention's programme theory by shining a light on the processes of change, such as perspectives on how and why the intervention leads to certain outcomes (Moore et al., 2015).

Quantitative methods, by contrast, were described as particularly appropriate for assessing measurable aspects of implementation theory such as reach (Proctor et al., 2011) or dosage (a key dimension of fidelity; Durlak & DuPre, 2008; Carroll et al., 2007), and the costs associated with implementation (Skivington et al., 2021). Quantitative approaches can also support examination of variation across sites or subgroups - an important aspect of understanding context (Moore et al., 2015; Craig et al., 2018). Combining quantitative indicators with qualitative insight strengthens the explanatory power of IPEs.

Standardised measures versus individual perceptions

A number of standardised validated measures of implementation constructs have been developed (see e.g. Aarons et al., 2014; Ehrhart et al., 2014; Weiner et al., 2017; Saldana et al., 2015; Mettert et al., 2020; Wensing, 2021), Overall, the literature suggests that evaluators should aim to use standardised measures of implementation wherever possible to strengthen comparability, rigour, and cumulative learning across studies.

While this point is most frequently articulated in relation to implementation outcomes, the same principle is equally relevant to programme theory – particularly in the assessment of intermediate outcomes or mechanisms (e.g. self-efficacy, confidence, motivation, wellbeing) that are hypothesised to link programme activities to longer-term outcomes. In principle, the use of validated, theory-linked instruments to assess such constructs can strengthen causal inference, reduce measurement error, and improve the interpretability of findings across contexts.

However, within the literature reviewed, there was limited explicit discussion of the use of standardised measures to assess programme mechanisms. Where intermediate outcomes were measured, authors more commonly relied on locally constructed or bespoke indicators tailored to the intervention context, rather than validated scales. This suggests a gap in current reporting and practice, rather than a lack of relevance, and points to an opportunity for future evaluations to more systematically align programme theory with established measurement tools where appropriate.

At the same time, locally constructed measures are sometimes necessary in evaluations of implementation and programme theory (Proctor et al., 2011; Wang et al., 2023) (Wolfenden et al., 2021), particularly where constructs are highly context-specific or where evaluations are conducted at early stages of intervention development (Cargo et al., 2018).

Authors have therefore called both for the development of a core set of standardised implementation outcome measures, and for improved reporting guidelines to ensure that researchers provide detailed descriptions of how measures – whether implementation-focused or programme-focused are defined, operationalised, and assessed (Wang et al., 2023).

Increase quality through consistency and shared standards of rigour

Finally, several authors highlight that testing and refining theory is strengthened using shared standards, consistent terminology, and validated frameworks that support analytical rigour. There is call for greater use of standardised and validated measures of implementation (Lengnick-Hall et al., 2023) and of frameworks such as the CFIR or RE-

AIM (Moullin et al., 2020) which identify key aspects of implementation and provide guidance on how to measure them. The complexity and heterogeneity of implementation contexts, the need to integrate multiple theories and the need for context-specific adaptation all raise challenges in the use of theories, models and frameworks (Nilsen (2015). However, they can inform the thinking of research teams by providing them with a shared language and practical tools for planning, conducting, and evaluating intervention implementation efforts, enhancing the viability and success of these efforts, and avoiding waste of resources and/or the derivation of erroneous conclusions (Moullin, et al., 2020).

While these arguments are most often articulated in relation to implementation theory, similar principles apply to programme theory. Programme theories are strengthened when they draw on shared conceptual definitions, use consistent terminology, and align measurement choices with clearly specified causal pathways. The use of established theories and validated constructs (e.g. self-efficacy, motivation, wellbeing) can support analytical rigour by reducing conceptual ambiguity. As with implementation theory, challenges arise from contextual heterogeneity and the need for adaptation; however, transparent specification remains critical to avoiding post hoc rationalisation and weak theorising.

Lengnick-Hall et al. (2023) discuss the need for "*shared standards of rigour*" and provide several practical recommendations for enhancing rigour in implementation research. They describe eight characteristics of high-quality implementation research to help provide a common language across evaluation teams:

1. Map and operationalise the specific multilevel context for defined populations and settings (e.g. individual, groups, organisations, systems)

2. Define and state the level of context for each construct under study

3. Describe how constructs relate to each other within and across levels

4. Specify the temporal scope of each phenomenon at each relevant level (i.e., determining the timing and duration of phenomena at each level to understand temporal dynamics)

5. Align measurement choices and construction of analytic variables with the levels of theories selected (and hypotheses generated, if applicable)

6. Use a sampling strategy consistent with the selected theories or research objectives and sufficiently large and variable to examine relationships at requisite levels

7. Align analytic approaches with the chosen theories (and hypotheses, if applicable), ensuring chosen techniques account for the nested structures inherent in multilevel data

8. Ensure inferences are made at the appropriate level of context

Although the above standards are articulated for implementation research, many of these characteristics are equally relevant to programme theory, particularly in relation to clarifying levels of analysis, specifying relationships between constructs, articulating temporal dynamics, and aligning measurement and analytic approaches with hypothesised causal mechanisms. There is a growing consensus that this heightened rigour will support more nuanced understandings of interventions and their implementation and therefore produce more meaningful learnings (Lengnick-Hall et al., 2023; Wang et al., 2023).

5. Equity in TIIPE

5.1. Introduction and definitions

Our database and grey literature searches identified very few papers that explicitly addressed equity in the design, evaluation or scaling of interventions and provided guidance for evaluators. As a result, we supplemented this chapter with additional texts from the implementation science literature that specifically focus on equity. This decision reflects the fact that, within the bodies of literature reviewed, equity is more consistently theorised and operationalised within implementation science than within programme theory. Consequently, we rely on a more limited pool of sources than other chapters in this review.

Implementation science is intrinsically connected to notions of equity (Brownson et al., 2021). Its emphasis on context, systems, and variation in delivery and uptake means that TIIPE is well positioned to explore whether and how interventions are effective for different populations, how implementation outcomes such as reach, adoption, acceptability and appropriateness vary across groups, and what this means for programme effectiveness and the need for differentiated content or implementation strategies - particularly for marginalised and underserved populations.

Equity considerations are also central to programme theory. Programme theories often contain implicit assumptions about who an intervention works for, how change occurs, and under what conditions. Without an explicit equity lens, these assumptions risk reflecting the experiences of more advantaged groups and obscuring variation in mechanisms, intermediate outcomes, and causal pathways across populations. Fletcher et al. (2016), for example, highlight the importance of examining how intervention mechanisms may operate differently across social groups, and how programme theory can either surface or conceal inequities depending on how it is specified and tested.

Incorporating an equity lens therefore requires attention to the needs, culture, and histories of key groups, alongside a critical analysis of systems and policies and how they operate for different populations to create inequitable outcomes. As Baumann & Cabassa (2020) note, *"The field of implementation science can help address these inequities by studying the factors, processes and strategies at multiple levels (e.g., clients, providers, organisations, communities) of the system of care that influence the uptake, use, and ultimately the sustainability of evidence-based interventions (EBIs), services and policies in community settings."*

Across fields, equity is commonly understood in terms of fairness and justice in the distribution of resources and opportunities so that equitable outcomes can be achieved (WHO, 2019). Within implementation science, this has typically involved understanding how to make interventions accessible, engaging, and beneficial to historically underserved populations, exploring how interventions can be adapted to meet the unique needs of different groups, and understanding the barriers and facilitators that may disproportionally exist for some groups and the implementation strategies required (Baumann & Cabassa, 2020). From a programme theory perspective, this also implies examining whether the hypothesised mechanisms of change are plausible, meaningful, and sufficient for different groups, and whether additional or alternative pathways may be required to achieve equitable outcomes.

The following sections outline how equity and TIIPE can align, and how to consider equity when conducting a TIIPE.

5.2. Alignment between equity and TIIPE

The relationship between equity and TIIPE extends to both the focus of the evaluation and the conduct of the research itself. Equity should be a key consideration within both the *design* and the *delivery* of evaluations, acknowledging that TIIPE is a method of evaluation that has significant potential to accelerate the achievement of equity goals by making visible how interventions operate across different contexts and populations (Baumann & Cabassa, 2020). These two distinct framings are mutually re-enforcing and have been referred to as a "*two-way street*" by Cabassa & Baumann (2013). As equity is increasingly considered in evaluation design, the barriers and facilitators to implementation for certain groups are more likely to be understood. And as group-level differences in barriers and facilitators are increasingly identified through rigorous evaluation, then interventions are more likely to be successfully refined to ensure equitable outcomes.

Brownson et al. (2021) recommend that every TIIPE project should include an equity focus. Although they acknowledge this may be more central in some evaluations than others, they caution that evaluators should still consider equity, e.g., when assessing implementation outcomes such as reach, to ensure that "*no one is left behind*", and that existing disparities are not "*inadvertently widened*" (Brownson et al., 2021). Fletcher et al. (2016) similarly emphasise the importance of realist approaches for exploration of how interventions work for different social groups and address inequalities in outcomes, and advocate for embracing systems-focused theory in logic models so that interventions respond effectively to inequality. They emphasise that feasibility and pilot studies should include a focus on implementation, feasibility and acceptability to the most deprived communities to support sustainable context-relevant solutions.

5.3. How to consider equity within TIIPE

5.3.1. Participatory approaches in evaluation processes and delivery

Commentators suggest that an important way to integrate equity into the design and delivery of TIIPE is to include a central role for lived experience. Cabassa & Baumann (2013) recommend using community-based participatory research (CBPR) methods, which involve communities in all stages of implementation and evaluation. This ensures that interventions are tailored to the specific needs and preferences of a community and that communities' voices are heard throughout the research process.

Participatory approaches are relevant not only to understanding implementation processes, but also to the development, testing and refinement of programme theory. Involving individuals with lived experience can help surface assumptions embedded within programme theories about how change is expected to occur, for whom, and under what conditions. From an equity perspective this is particularly important because dominant programme theories may reflect the experiences and priorities of more advantaged groups, while overlooking alternative mechanisms or outcomes that are meaningful for marginalised populations (Fletcher et al., 2016).

Participatory approaches are therefore not only seen as a way of realising equity within TIIPE, but also as a means of strengthening the quality and relevance of evidence (Frye and Hemmer, 2012; Seward et al., 2021).

The following examples from Minkler, Salvatore & Chang (2017) demonstrate how participatory approaches can be woven into TIIPE to promote greater equity:

- Co-producing overarching research questions to ensure they focus on relevant issues to all groups being offered the intervention
- Co-designing data collection tools to allow interviews to be a safe space for participants and practitioners to reflect on issues that are most closely connected to their experiences
- Co-producing the focus for analysis and interpretation of data may help to mitigate subjective biases of research teams and continue to anchor evidence led recommendations in the experiences of marginalised communities
- Promoting participant voice in the dissemination of learnings can support policy makers and practitioners to contextualise TIIPE learnings in the context of lived experience

5.3.2. Equity-specific outcomes, mechanisms and frameworks

Addressing equity in TIIPE involves assessing how effectively and why interventions are reaching and benefiting different population groups, especially those who are marginalised or disadvantaged, and how implementation and programme effectiveness can be increased. Researchers need to systematically evaluate how contextual factors (e.g., socioeconomic status, ethnicity, geography) impact both the implementation processes, but also the operation of programme mechanisms and outcomes (Cullen & Walsh, 2020).

From a programme theory perspective, this involves interrogating whether hypothesised mechanisms of change are plausible and meaningful for different groups, whether they operate in the same way across contexts, and whether additional or alternative mechanisms may be required to achieve equitable outcomes. Fletcher et al. (2016), for example, emphasise the importance of examining how interventions interact with wider social systems and how mechanisms may be enabled or constrained differently across social groups, with implications for both programme design and evaluation.

From an implementation theory perspective, one approach to considering equity within TIIPE is to develop equity-specific indicators for key implementation outcomes (such as reach, acceptability, adoption, fidelity, and sustainability) and track whether more disadvantaged groups are benefiting as much as other groups (Baumann & Cabassa, 2020), for example:

- **Reach** – an examination of equity in TIIPE requires thoughtful attention to reach and representation (Baumann & Cabassa, 2020). Evaluators could explore questions such as who is participating and engaged and who is missing from implementation studies? Understanding these metrics can help identify if certain groups face greater barriers to participation and if diversified strategies are needed to improve their engagement (Glasgow et al., 2019).
- **Appropriateness and adaptation** – these outcomes could be applied to address equity by exploring whether interventions are viewed as appropriate by all groups and/or individuals, and if they have been adapted for equity, e.g., whether adjustments were made to ensure programme content and approaches are culturally appropriate (Brownson et al., 2021). Commentators suggest that adaptation of interventions should go beyond linguistic translation to include cultural tailoring, which ensures that content, delivery methods, and the overall approach are relevant to the specific cultural context of the target population (Stirman et al., 2013; Baumann & Cabassa, 2020).
- **Fidelity** - evaluators could attempt to track intervention fidelity in different contexts, to see whether interventions are being delivered consistently across diverse groups

and in different settings, and if not, explore how adaptations could help to improve equity (Stirman et al., 2013)

Specific equity-focused implementation science frameworks have also been developed, including the Health Equity Implementation Framework (Woodward et al., 2021). The Health Equity Implementation Framework proposes determinants of inequitable implementation and a process (facilitation) by which to address determinants (Woodward et al., 2021). The domains contained in this framework are those known to affect health disparities, and include culturally relevant factors, the client-provider interaction, and the societal context (Palmer et al., 2019).

5.3.3. Disaggregation of data

Assessing equity within both programme theory and implementation theory requires disaggregation of data by relevant demographic and contextual variables, where sample sizes allow (e.g., age, ethnicity, gender, income, geographic location). This enables researchers to track whether certain groups are experiencing different outcomes or impacts compared to others, helping to identify areas where equity gaps may exist (Baumann & Cabassa, 2020). If differences are found, commentators suggest that evaluators should assess how programme design, hypothesised mechanisms, and implementation strategies can be refined or adapted to reduce these disparities and create more equitable outcomes (Baumann & Cabassa, 2020).

6. Incorporating theory at different evaluation stages

This chapter focuses on RQ3: *How should evaluators design evaluations to test programme theory and implementation theory, at different stages of evaluation and in different contexts?*

6.1. Youth Futures evaluation stages

Youth Futures commission a range of projects, corresponding to different evaluation stages:

- **Intervention design stage**: Youth Futures commissions developmental evaluations to support the design of a programme, the development of a clearly defined theory of change, and to build capacity for future evaluation.
- **Feasibility stage**: Youth Futures commissions feasibility studies to explore whether the conditions are in place for a programme to be evaluated through an impact evaluation, and whether any refinements or adaptations need to be implemented to enable the programme to be evaluated through an impact evaluation.
- **Pilot stage**: Youth Futures commissions pilot studies to identify whether a proposed impact evaluation design will work in practice, and whether any refinements need to be made ahead of launching a full impact evaluation. This may involve piloting the design at a smaller scale, or piloting specific elements of the design (e.g. randomisation procedures).
- **Impact stage**: At this stage Youth Futures commissions the evaluation of a programme through a randomised controlled trial (RCT), quasi-experimental design (QED), or alternative robust method.

Stages may occur in parallel e.g. the intervention design stage may be commissioned alongside a feasibility stage, or a feasibility study may be commissioned in combination with a pilot study.

6.2. TIIPE considerations at different evaluation stages

Although the reviewed texts generally did not discuss evaluation stages in ways that aligned precisely with the Youth Futures study types, the consensus is that TIIPE can and

should be conducted across the intervention development and evaluation spectrum. TIIPE is relevant at all evaluation stages, although the focus of TIIPE is likely to vary by stage (Humphrey et al. 2016a, b; Moore, et al., 2015).

Integrating equity considerations throughout the various stages of programme evaluation is also essential. Recommendations below summarise how to embed equity at each evaluation phase and draw on data from a review of participatory approaches commissioned by Youth Futures (Rowland et al., 2024), as well as other relevant literature (e.g., Minkler, Salvatore & Chang, 2017).

6.2.1. Intervention design stage

From a programme theory perspective

At the intervention design stage, the focus of IPE is on optimising programme design and laying the foundations for further evaluation. From a programme theory perspective, this typically involves developing or strengthening the programme's ToC, identifying hypothesised causal pathways, assumptions and contextual influences (Moore et al., 2015; Skivington et al., 2021; Humphrey et al., 2016; Edmunds et al., 2022). Formative or pre-evaluation research may be necessary to support this process, including exploring potential mechanisms of change and refining delivery approaches in light of early learning (Cargo et al., 2018; Moore et al., 2016; Fletcher et al., 2016). Developing a well-specified and testable ToC, is a key output of this phase and provides the foundation for subsequent feasibility and pilot testing.

From an implementation theory perspective

The intervention design stage also marks the point at which evaluators begin to articulate an initial implementation theory. This includes identifying anticipated implementation determinants, early barriers and enablers to delivery, and hypotheses about which implementation strategies may be needed to support uptake and fidelity in different contexts (Skivington et al., 2021). At this stage, evaluators and implementers typically begin shaping the overall evaluation design and establishing monitoring and evaluation (M&E) infrastructure to track early implementation outcomes (e.g., reach, fidelity, acceptability). Building organisational capacity and data systems at this point supports more systematic testing and adaption of implementation assumptions during later stages of evaluation.

From an equity perspective

Equity considerations should be explicitly embedded during intervention design, as decisions made at this stage shape whose needs and experiences are reflected in both

programme and implementation theory. Key actions include engaging young people from marginalised backgrounds in defining and refining the core elements of a programme theory so that it reflects lived experience and addresses the specific challenges these populations face (Rowland et al., 2024). The intervention design stage is also a critical moment to establish inclusive practices for co-design and supportive environments that enable all voices to be heard and valued (Rowland et al., 2024). This may include collaborative development of data-collection tools or identifying creative and culturally appropriate methods to ensure that future evaluation activities meaningfully capture diverse participant experiences (Rowland et al., 2024).

6.2.2. Feasibility stage

From a programme theory perspective

The feasibility stage is typically the first point at which a programme is delivered in practice, providing an initial opportunity to assess whether the programme theory (often articulated through the ToC) holds. A central task at this stage is to determine whether the programme has reached a degree of stability before more formal evaluation begins. Durlak and Dupre (2008) argue that no programme should be evaluated until sufficient time has been allowed for its implementation to stabilise, as premature evaluation risks producing misleading data about feasibility or effectiveness.

Testing programme theory at this stage includes examining whether the programme components are delivered, whether hypothesised mechanisms appear plausible in practice, whether contextual assumptions are borne out, whether the intended outcomes are observed and are measurable for later evaluation stages (Moore et al., 2015). Feasibility studies provide an opportunity to refine the ToC by exploring emerging mechanisms and gaining a more grounded understanding of how context shapes delivery. They also support early decisions about what aspects of the programme theory require further clarification and which outcomes are appropriate to be assessed in a later impact evaluation. Although programme theory is a central focus at this stage, refinement continues to be iterative and ongoing across subsequent evaluation stages.

From an implementation theory perspective

While the intervention design stage may have identified anticipated barriers and enablers, and implementation strategies – feasibility studies provide the first opportunity to examine these issues systematically using data from real-world delivery. At this stage, evaluators assess whether the intervention can be implemented as intended within the available organisational and resource context, and whether its core components can be maintained during initial delivery. Even when a programme is being implemented for the first time, feasibility studies allow evaluators to explore early fidelity and adaptation

patterns and to identify which implementation strategies may require adjustment before piloting or larger-scale testing. This stage therefore contributes to the development of implementation theory by generating initial evidence about how the intervention interacts with context and what supports or threatens implementation success.

From an equity perspective

Feasibility is the first point in the evaluation pipeline to examine systematically how different groups experience and access the programme in practice (Proctor et al., 2011; Powell et al., 2015). This includes examining whether the programme is more acceptable to some groups than others and whether adaptations may be required to ensure relevance and accessibility across diverse populations. Early feasibility work can also test approaches that facilitate broader participation, such as refined recruitment strategies or more inclusive delivery methods (Powell et al., 2015).

6.2.3. Pilot stage

From a programme theory perspective

The pilot stage provides an opportunity to further test and refine the programme theory before moving into an impact evaluation. Authors emphasise the importance of using pilot work to examine whether the hypothesised pathways in the ToC appear plausible in practice, and whether the constructs and outcomes identified are the right ones to measure at scale (Humphrey et al., 2016a and b; Moore et al., 2015). This is the stage when moderators and intermediate outcomes identified in the programme theory would be tested (Humphrey et al., 2016a and b; Moore et al., 2015). Fidelity assessments conducted during pilot delivery also support programme theory refinement by showing whether core components can be delivered as intended (Proctor et al., 2011). This stage therefore contributes to strengthening the clarity, validity and measurability of the programme theory ahead of a trial.

From an implementation theory perspective

Pilot studies play a key role in testing implementation strategies and understanding how these may need adapting—including for different populations—to improve implementation quality and outcomes (Seward et al., 2021). Proctor et al. (2011) highlight adoption, feasibility, and appropriateness as important considerations during this stage, helping evaluators assess whether the intervention fits the delivery context and whether modifications are required to enhance acceptability and uptake. Fidelity assessment remains critical to ensure that adaptations do not compromise the core components of the intervention (Proctor et al., 2011). Pilot studies are also a suitable

point to begin assessing cost to understand the financial feasibility of future scale-up (Proctor et al., 2011).

From an equity perspective

Equity considerations at the pilot stage involve assessing whether implementation strategies are working effectively for all intended populations and whether adaptations may be required to improve accessibility and engagement. Participatory approaches, including involving young people as co-evaluators, can help ensure that marginalised groups' perspectives shape the evaluation process and inform decisions about programme refinements (Rowland et al., 2024). Attention to the experiences of different groups at this stage supports more inclusive and responsive implementation planning and programme refinement before moving to an impact trial.

6.2.4. Impact stage

From a programme theory perspective

At the impact stage, programme theory continues to play an important role in interpreting trial findings and understanding why an intervention does or does not produce its intended effects. TIIPE helps contextualise impact results by analysing the quality and consistency of delivery, identifying differential effects across groups, and examining whether the hypothesised pathways in the ToC hold in practice (Humphrey et al., 2016a and b; Moore et al., 2015; Skivington et al., 2021). This is also the stage when moderators and intermediate outcomes can be tested with larger samples. Because complex interventions may evolve as they are scaled, ongoing programme theory refinement is still needed at this stage, drawing on new insights about emerging barriers, delivery challenges or unexpected outcomes (De Silva et al., 2014; Fletcher et al., 2016). Several authors emphasise that theory refinement remains important throughout all evaluation phases, including impact and scale-up (Humphrey et al., 2016; Skivington et al., 2021).

From an implementation theory perspective

During the impact phase, TIIPE typically focuses on assessing implementation quality at scale, including fidelity, reach and variation in delivery across sites, to understand how implementation influences overall effectiveness (Moore et al., 2015; Proctor et al., 2011). This includes examining whether certain implementation barriers or facilitators help to explain variation in outcomes and whether additional support or refinement of implementation strategies is needed. As interventions move into later stages of delivery, assessing sustainability and penetration becomes increasingly important, including understanding whether the intervention becomes embedded within organisational

practice, supported through policy or maintained through ongoing funding (Proctor et al., 2011; Moore et al., 2015). Because new implementation challenges often emerge at scale, a full TIIPE may still be needed to support interpretation and guide adaptation (Moore et al., 2015).

Hybrid implementation–effectiveness studies provide one mechanism for deepening this form of integrated testing. Hybrid designs allow evaluators to examine effectiveness and implementation simultaneously (Curran et al., 2012; 2022). Hybrid type 1 studies focus on effectiveness while also collecting some implementation data. Hybrid type 2 studies give equal emphasis to effectiveness and implementation, enabling a more rigorous test of implementation theory in real-world settings and its relationship with programme effectiveness. Hybrid type 3 studies focus primarily on implementation, testing different implementation strategies against each other while also monitoring outcomes e.g. through routine data. Hybrid type 2 and 3 designs allow evaluators to explore how variation in implementation, such as levels of fidelity or the use of specific strategies, is associated with programme impact, strengthening understanding of how implementation processes shape outcomes at scale.

From an equity perspective

Equity remains a core concern at the impact stage, particularly in understanding whether benefits are experienced evenly across groups. Disaggregated analyses can help identify whether some populations experience weaker impacts or face persistent implementation barriers, informing adjustments to programme content and delivery to improve equitable outcomes (Wolfenden et al., 2021). TIIPE contributes by identifying whether all groups within the target population are reached, whether they experienced the intervention differently, whether fidelity varied for populations, and whether improvements in programme content and implementation are needed.

7. The use of theory in evaluations of interventions to support the employment of marginalised young people

To address RQ4 (*How has TIIPE been undertaken in evaluations of interventions to support the employment of marginalised young people*?), we reviewed evaluations of interventions published by Youth Futures and studies noted in the Youth Futures Evidence and Gap Map (EGM), assessing whether and how programme and implementation theory were incorporated.

7.1. Use of programme theory in the selected evaluations

Where programme theory was articulated in the evaluations reviewed, this was consistently presented in the form of a ToC, typically described using a combination of visual diagrams and narrative explanation. The level of detail and analytical depth varied considerably across reports. Several Youth Futures-commissioned evaluations explicitly described the development and testing of ToCs (Adams et al., 2023; Beninger et al., 2024; Marren et al., 2024). In contrast, explicit reference to programme theory or the use of a ToC was absent in several of the studies identified from the EGM.

7.1.1 Presenting programme theory

The ToCs in the evaluation reports reviewed included the following elements:

- Context and rationale
- Assumptions
- Activities
- Inputs
- Outputs
- Outcomes:

- short-, medium- and long-term
- specific quantifiable outcomes, wider outcomes

- Impacts

Some features of ToCs were notably absent in these reports, including a description of the target group and the mechanisms of change. There was also variation in how and how far causal chains or causal pathways were articulated.

As noted in chapter 5, programme theories and ToCs are strengthened when they are specified across multiple levels and when they explicitly represent the implementation approach alongside causal pathways. However, in the evaluations reviewed, ToCs generally focused on causal pathways and outcomes for programme recipients only, with limited attention to wider levels such as delivery staff, organisations or systems. While this focus may reflect the primary aims of these evaluations, it means that causal pathways and outcomes relating to delivery staff and organisations were often not articulated. These elements can provide important context for understanding how programmes are intended to operate in practice and how implementation processes interact with programme mechanisms. Their absence therefore suggests missed opportunities to articulate and test hypotheses about organisational- and system-level factors and outcomes, which could have supported a more comprehensive and theory-informed understanding of programme operation and potential impacts.

7.1.2 Theory of change development and refinement

Across the evaluations reviewed, the ToC was typically developed through a combination of tacit knowledge and structured input from several stakeholders, including delivery teams and evaluators. In some cases, this work involved refining or formalising an existing, implicit programme theory, rather than developing a ToC entirely from scratch - an approach that is not always made explicit in evaluation reporting.

Atkinson et al. (2017), for example, provide a detailed description of how an existing programme theory was surfaced and refined through participatory workshops. They specified who was involved, how contributors were supported to engage in the process, and how outputs from these workshops were triangulated with documentary evidence to produce a more explicit and testable ToC for use in a subsequent impact evaluation. Other studies also referred to situating or refining their programme logic in relation to the wider evidence base, for example by linking core features of the intervention model to existing research or theory (Johnson et al., 2023; Beninger et al., 2023).

While these examples demonstrate good practice in articulating and refining programme theory, explicit discussion of ToC refinement, as distinct from initial

development, was relatively rare across the evaluations reviewed. This suggests that opportunities to make theory refinement more transparent, and to document how programme theories evolve over time, may be underutilised in evaluation reporting.

7.1.3. Testing programme theory

Testing programme theory was a central function across the evaluations reviewed, with several studies demonstrating that initial ToCs did not fully reflect delivery in practice. In some cases, this was due to substantial contextual disruption, such as changes required during the COVID-19 pandemic (Mackay et al. 2023), while in others it reflected misalignment between the anticipated and actual appropriateness of the model for the participant group (Beninger et al. 2023). These findings underline the importance of designing IPEs that explicitly test programme theory and allow for systematic refinement in response to emerging evidence.

To extent to which programme theory testing and revision were described varied across the reports. Beninger et al. (2023), for example, clearly described the elements of the initial ToC, based on two pre-existing programme logics and wider literature, and described this was revised as evidence emerged about the programme's causal pathways. Mackay et al. (2023) similarly reported that the original ToC required updating because changes to delivery conditions and participant engagement meant that some of the original assumptions about how and why the programme would work no longer held. In response, the ToC was revised in collaboration with delivery staff to reflect adapted delivery models, revised mechanisms of change, and updated assumptions about contextual influences. These updates were explicitly documented at the end of the evaluation.

Together, Beninger et al. (2023) and Mackay et al. (2023) demonstrate the importance of a clear and evidence-informed programme theory that can be systematically and transparently investigated through an IPE. Both studies provide detailed accounts of how assumptions underpinning the ToC were examined and whether they held in practice, and how findings informed refinements to programme design, including eligibility criteria, delivery modality, and session timing and duration. In both cases, mechanisms of change were identified empirically as findings of the evaluation, rather than being fully specified in advance and formally tested, highlighting both the strengths and limitations of current approaches to programme theory testing in practice.

7.2. Use of implementation theory in the selected evaluations

Across the evaluations reviewed, none explicitly articulated an implementation theory or framed their analysis using a named implementation framework. However, many studies examined aspects of implementation in practice, often implicitly using non-standardised terminology. Commonly recognised implementation constructs such as reach, acceptability, appropriateness, feasibility were explored but were not consistently labelled or organised as part of a coherent implementation theory. For example, Mackay et al. (2023) explored how contextual disruption affected delivery and participant engagement but did not frame this analysis using implementation science constructs or terminology. Similarly, some evaluations appeared primarily focused on estimating impact (e.g. Brandenburg et al., 2016) or refining programme theory (e.g. Adams et al., 2023), which may partly explain the absence of explicit implementation theory despite the presence of relevant implementation-related analysis.

Many of the included evaluation reports – particularly those funded by Youth Futures – nonetheless engaged substantively with implementation-related research questions. For example, Beninger et al. (2023) explored how contextual factors influenced implementation and how these shaped participants' experiences. Beninger et al. (2023) and Johnson et al. (2023) both explored whether the intended delivery model could be implemented in practice, describing this in terms of the *plausibility of the ToC* rather than feasibility per se.

Across the reviewed reports, evaluators also examined how appealing programmes were to participants and practitioners (acceptability), whether adaptations were required to ensure suitability for participants (appropriateness), and the effectiveness of referral routes and eligibility criteria in shaping participation (reach). However, these implementation dimensions were covered inconsistently and described using varied and study-specific language. As a result, opportunities to make implementation assumptions explicit, to articulate a coherent implementation theory, or to build cumulative evidence about implementation factors across commissioned studies appear to have been missed.

8. Discussion

8.1. Summary

This report builds on recent guidance developed by several What Works Centres and central Government and extends this work to include a more central focus on theory within evaluation, covering both programme and implementation theory.

This report documents findings from a rapid evidence review of published journal articles and grey literature, focused on texts which offered guidance on how to use programme theory and/or implementation theory within IPE and evaluation more generally. The search approach was pragmatic, including a search of specific academic databases, a targeted website search, and expert consultation. Papers were included if they were published between the year 2000 – May 2024, and covered aspects of human services, with no geographic exclusion. Fifty texts were selected and extracted, including guidance on designing, testing and refining programme theory; guidance on assessing implementation; implementation frameworks; and guidance on equity-focused IPEs.

8.2. The role of theory in IPE

Theory-informed implementation and process evaluation (TIIPE) is vital for understanding how and why interventions succeed or fail in different contexts.

The integration of theory into IPE is fundamental for several reasons. Using theory can provide a roadmap for understanding the mechanisms through which interventions work, highlighting critical factors that influence their success. A well-defined programme theory can guide evaluators in assessing core components of an intervention and their links to outcomes. A strong theoretical foundation helps evaluators identify moderators and mediators - contextual or individual-level variables that may impact the intervention's outcomes (Proctor et al., 2023). TIIPE draws on theories about behavioural change and the implementation strategies required to catalyse it, and the use of theoretical frameworks is critical in guiding evaluators to make sense of complex, real-world settings (Proctor et al., 2023).

For example, frameworks such as the Consolidated Framework for Implementation Research (CFIR) help evaluators to understand how determinants of implementation across different ecological levels can influence outcomes (Damschroder et al., 2009 and 2022). Similarly, theories of behaviour change, such as the Theoretical Domains Framework (TDF) can guide evaluators in examining multi-level factors that may affect

the implementation of interventions (Michie et al., 2009). By linking theoretical constructs to specific implementation strategies, evaluators can better understand why certain outcomes occur and how to improve programme design and implementation efforts.

Using theory also supports transparency and replicability which can help advance the evidence base and close the research-practice gap (Sekhon et al., 2017; Nilsen, 2015; Lafferty & Mahoney, 2003). By producing evidence that is more useable by and useful to decision makers (Skivington, et al., 2021), it can also lead to increased impact on policy and public understanding (Lafferty & Mahoney, 2003).

8.3. Challenges in integrating theory in IPE

Despite the central role of theory in IPE, several challenges arise in applying programme and implementation theories in practice. These challenges do not reflect limitations of theories, models, or frameworks (TMFs) per se, but rather the practical and analytical demands of using them effectively in complex, real-world evaluations.

A key challenge relates to how theory is applied in settings characterised by contextual complexity and variation. Many interventions are implemented across diverse organisational, system, and cultural contexts, and evaluators must account for how these differences shape both implementation and outcomes. While TMFs can help structure thinking about key constructs and influences, applying them across heterogeneous settings often requires drawing on multiple theories or frameworks to capture different aspects of context and change processes (Nilsen, 2015).

A further challenge concerns the practical use of TMFs to support theory-driven analysis, particularly when seeking to understand mechanisms. Although the growing number of TMFs has expanded, these frameworks are not always sufficient on their own to support in-depth analysis of *how* implementation processes unfold in practice. TMFs often provide limited guidance on how contextual factors, implementation strategies, and programme components interact dynamically to generate outcomes (Nilsen, 2015; Proctor et al., 2023).

This has led some scholars to criticise TMFs as describing the '*anatomy*' of implementation but failing to address the '*physiology*', i.e., how implementation functions and contributes to outcomes (Davidoff, 2017). This distinction is especially salient for IPE, which seeks to explain causal processes rather than simply catalogue implementation conditions. As a result, evaluators are often required to move beyond the descriptive use of TMFs and engage in more explicit, prospective theorising about mechanisms linking programme components, contextual influences, implementation

strategies, and outcomes. The systematic testing and refinement of these hypothesised mechanisms is necessary to advance both programme and implementation theory.

A further challenge relates to equity in TIIPE and applies to both programme theory and implementation theory. The evolving nature of theory development in these fields means that some theories and frameworks may still be under development or have not yet been tested across diverse populations and settings, creating uncertainty about how they should be adapted or applied within specific evaluations (Baumann & Cabassa, 2020).

8.4 Summary of key themes around TIIPE guidance

Several strategies can support evaluators in addressing these challenges and integrating theory effectively within IPE. A summary of the key themes for TIIPE guidance is outlined below.

8.4.1. Develop and test programme and implementation theory iteratively and collaboratively

Theory development should be understood as an iterative process, involving cycles of theory building, testing, review and refinement. This process typically draws on both formal evidence and tacit knowledge and is often collaborative, involving delivery teams, evaluators and other stakeholders. Although there is broad consensus that testing and refining theory is a key aspect of TIIPE, there is currently limited practical guidance on how to operationalise this in practice – for example, how much and what type of evidence is sufficient to confirm, challenge, or require change to, programme or implementation theory. Clearer guidance in this area could support more transparent, consistent theory-informed evaluation practice.

8.4.2. Explore and account for context

Throughout the design and conduct of TIIPE, a crucial recommendation is to pay close attention to context (Cargo et al., 2018). Evaluators should always begin by considering the context in which the intervention is being implemented. Frameworks such as CFIR (Damschroder et al., 2009 and 2022) or Promoting Action on Research Implementation in Health Services (PARIHS) (Harvey & Kitson, 2016) are particularly useful in understanding how contextual factors (e.g., organisational culture, leadership support) influence implementation. For community-based interventions, community-oriented

frameworks (e.g., the Community-Based Participatory Research (CBPR) model) (Cabassa & Baumann, 2013) may be more appropriate.

8.4.3. Incorporate multiple theories / theoretical approaches

As no single theory can fully explain the complexity of a programme and its implementation, evaluators should also consider using multiple theories in TIIPE (Moore, et al., 2015; Nilsen, 2015).

A rapidly growing body of studies use multiple theoretical approaches to test and refine programme theory, e.g., integrating realist evaluation approaches with theory of change approaches to develop a more mechanism-driven and context-sensitive programme theory (Rolfe, 2019).

Implementation theories have also been combined with behavioural theories to more thoroughly explore the organisational and individual factors that impact implementation. This approach can help clarify how internal (e.g., staff motivation) and external (e.g., community support) factors interact to influence outcomes (Nilsen, 2015).

8.4.4. Flexible approach to theory refinement

Given the dynamic nature of real-world implementation, both programme and implementation theories should be treated as provisional and open to refinement within a TIIPE. Ringhofer & Kohlweg (2019) argue that programme theories must be flexible to be useful, in the sense that they can be revised when underlying assumptions or mechanisms, for example, do not hold in practice. As new data emerge, evaluators may also need to adapt the focus, methods, or analytic priorities of the evaluation to test revised hypotheses or explore newly identified mechanisms. Flexibility in theory refinement is distinct from but closely linked to flexibility in the design and conduct of the TIIPE itself. This adaptive approach allows evaluators to both refine programme and implementation theories *and* adjust the TIIPE so that it remains aligned with how the intervention is unfolding in practice (De Silva et al., 2014).

8.4.5 Using standardised measures

Assessing programme and implementation theory requires evaluators to translate theoretical constructs into observable and measurable indicators. This applies not only to implementation outcomes (e.g., reach, fidelity, acceptability), but also to key elements of programme theory, including hypothesised mechanisms of change, intermediate outcomes and relevant aspects of context. While the most appropriate measures will vary by programme model, population, and setting, and the development of standardised validated instruments remains ongoing, many authors argue that existing

measures are under-used, making it difficult to compare findings across studies and populations (Proctor et al., 2011; Wang et al., 2023).

In response, there have been calls for greater use of validated measures where appropriate, alongside further development and validation of instruments that capture key constructs of programme and implementation theory. Authors also emphasise the importance of improved reporting practices, with clearer documentation of how constructs are defined, operationalised, and assessed – regardless of whether standardised or locally developed measures are used (Wang et al., 2023). Together, these advances could strengthen the rigour, transparency, and interpretability of TIIPEs.

8.4.6 Adjust the focus of TIIPE at different evaluation stages

There is a consensus that TIIPE is relevant at all evaluation stages, although the focus and questions addressed will change. In pre-evaluation research and early evaluation, the focus of IPE is on developing, testing and refining the programme theory and assessing early implementation issues such as acceptability, feasibility and barriers to implementation. Later the focus shifts to contextualising impact findings, explaining variation, testing aspects of programme theory such as mechanisms and moderators, exploring associations between implementation and effectiveness, and testing implementation strategies against each other.

8.4.7. Essential role for stakeholder engagement and equity

Incorporating stakeholder perspectives into the evaluation design is essential for ensuring that TIIPE is relevant and culturally appropriate (Minkler, Salvatore & Chang, 2017). Designing a TIIPE study to capture views from multiple groups connected to an intervention (e.g. developers, delivery staff, clients, community leaders) can also help to identify contextual factors that may not be captured by existing theories (Cabassa & Baumann, 2013). Throughout the stages of TIIPE, it will be important to explore whether the programme design, content and implementation approaches are at least as appropriate for marginalised groups as for others. There should also be a focus on equity in dissemination efforts. Involving stakeholders in the analysis and interpretation of data can help to ensure evaluation findings are expressed in ways that are meaningful to those directly involved and impacted by the intervention (Brownson et al., 2021).

8.4.8. Dissemination of theory-driven findings

Centring theory in the reporting of findings is crucial for advancing the field (Brownson et al., 2013). Evaluators should clearly articulate the original programme and implementation theory, how this was operationalised and tested, how programme and implementation theory were refined, and how this focus contributed to their

understanding of the programme and its implementation. More consistency in language and the constructs on which evaluation is centred will support a richer understanding of how programmes need to be designed and implemented to best support marginalised young people's employment outcomes. Sharing this information in dissemination efforts will help to build a stronger body of evidence for theory development and its application in future studies (Proctor et al, 2023). By disseminating theory-driven findings, evaluators contribute to a continuous feedback loop where practice-based evidence can inform theory refinement and where theory refinement leads to more effective programmes and implementation.

8.5. Strengths and limitations of the review

8.5.1 Strengths

This is, to our knowledge, the first effort to summarise existing guidance on TIIPE with an explicit focus on programme and implementation theory. Previous reviews have summarised evidence on aspects of TIIPE, such as how to develop programme theories of change, conduct IPEs, or use theory-based approaches in evaluation. As such the review builds out from guidance developed by UK What Works Centres and central government.

8.5.2 Limitations

A key limitation in the review is that the term 'TIIPE' was not used in the texts reviewed. We therefore had to construct our own definition of TIIPE a priori, as combining both programme and implementation theory. The literature reviewed generally approached programme theory through discussions of the use of ToCs, with relatively little reference to higher level theories informing programme design. Similarly, the literature on implementation discussed key implementation constructs which can be seen as theoretical approaches (Nilsen, 2015) but are not, in themselves, necessarily theories about implementation.

We were aware that it would be challenging to scope and summarise the extensive literature within the available time and resources. Our search and selection strategy focused on guidance documents relevant to how either programme theory or implementation theory or (explicitly or implicitly) both should be considered in IPE. However, when we synthesised the extracted evidence, there were gaps in the intended coverage of the review, which we filled with pragmatic and non-systematic selection of further texts known to the review team. This extended the review coverage, but introduces bias, and we were conscious still of relying on a small number of selected texts in addressing some issues.

Several gaps remain. As we expected, we found no guidance on TIIPE specifically relevant to Youth Futures' focus on interventions to support marginalised young people into employment. Although there was extensive writing on how to develop programme theory, we found little specific guidance on how to interpret data to validate or refine aspects of theory - and how to judge whether evidence is sufficient or not to consider an aspect of theory (e.g. a mechanism or causal link) proven. We found limited discussion of equity and supplemented this chapter with several additional texts.

The individual evaluation reports we reviewed tended to focus discussion of programme theory around the development and testing ToCs and often did not provide rich or nuanced accounts of theory through the lens of either programme or implementation theory.

8.6. Conclusion

TIIPE is an essential tool for understanding and improving the effectiveness of interventions in real-world settings. By grounding evaluations in programme and implementation theory, evaluators can better understand the factors driving implementation success (or failure) and identify critical factors that influence outcomes and affect transferability. While building theory into IPE is challenging given the complex real-world settings of interventions and their evaluation, adopting a flexible, context-specific approach, combined with an equity focus and meaningful stakeholder engagement, and aiming for explicit operationalisation of theory can ensure that TIIPEs are both rigorous and relevant. As the field continues to evolve, advancing methods together will strengthen the capacity of evaluators to build learning about effective programme development and implementation.

References

Texts marked with an asterisk were added after the initial literature search.

*Aarons, G.A., Hurlburt, M., Horwitz, S.M. (2011). Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health and Mental Health Services Research*, *38*, 4-23.

*Aarons, G.A., Ehrhart, M.G., & Farahnak, L.R. (2014). The Implementation Scale (ILS): development of a brief measure of unit level implementation leadership. *Implementation Science : IS*, 9(1), 45. https://doi.org/10.1186/1748-5908-9-45

Albers, B., Mildon, R., Lyon, A.R., Shlonsky, A. (2017). Implementation frameworks in child, youth and family services – Results from a scoping review. *Children and Youth Services Review*, 81, pp.101–116. https://doi.org/10.1016/j.childyouth.2017.07.003

*Astbury, B., & Leeuw, F. L. (2010). Unpacking black boxes: mechanisms and theory building in evaluation. *American Journal of Evaluation, 31*(3), 363-381.

*Baumann, A.A., & Cabassa, L.J. (2020). Reframing implementation science to address inequities in healthcare delivery. *BMC Health Services Research*, *20*, 1-9.

*Blamey, A., and Mackenzie, M (2007) *Theories of Change and Realistic Evaluation. Peas in a Pod or Apples and Oranges? Evaluation* SAGE Publications. Vol 13(4): 439-455. https://doi.org/10.1177/1356389007082129

*Breuer, E., Lee, L., De Silva, M. and Lund, C. (2015). Using Theory of Change to Design and Evaluate Public Health Interventions: a Systematic Review. *Implementation Science*, 11(1). https://doi.org/10.1186/s13012-016-0422-6.

*Brownson, R.C., Jacobs, J.A., Tabak, R.G., Hoehner, C.M., Stamatakis, K.A. (2013). Designing for dissemination among public health researchers: findings from a national survey in the United States. *American Journal of Public Health*, 103(9):1693–9. https://doi.org/10.2105/AJPH.2012.301165

*Brownson, R.C., Kumanyika, S.K., Kreuter, M.W., & Haire-Joshu, D. (2021). Implementation science should give higher priority to health equity. *Implementation Science*, *16*, 1-16.

*Cabassa, L.J., & Baumann, A.A. (2013) A two-way street: Bridging implementation science and cultural adaptations of mental health treatments. *Implementation Science,* 8:90.

Cargo, M., Harris, J., Pantoja, T., Booth, A., Harden, A., Hannes, K., Thomas, J., Flemming, K., Garside, R. and Noyes, J. (2018). Cochrane Qualitative and Implementation Methods Group guidance series—paper 4: methods for assessing evidence on intervention implementation. *Journal of Clinical Epidemiology*, 97, pp.59–69. https://doi.org/10.1016/j.jclinepi.2017.11.028

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2(1). https://doi.org/10.1186/1748-5908-2-40

*Chen, H.T. (2012). *Theory-driven evaluation: Conceptual framework, application and advancement* (pp. 17-40). Springer Fachmedien Wiesbaden.

*Clarke, A.R., Goddu, A.P., Nocon, R.S., Stock, N.W., Chyr, L.C., Akuoko, J.A., Chin, M.H. (2013) Thirty years of disparities intervention research: what are we doing to close racial and ethnic gaps in health care? *Medical Care*, 51(11):1020–6. https://doi.org/10.1097/MLR.0b013e3182a97ba3

*Collins, L.M., Murphy, S.A. and Strecher, V. (2007). The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. *American Journal of Preventive Medicine*, 32(5), pp.S112-S118.

Craig, P., Di Ruggiero, E., Frohlich, K.L., Mykhalovskiy, E., White, M., Campbell, R., Cummins, S., Edwards, N., Hunt, K., Kee, F., Loppie, C., Moore, L., Ogilvie, D., Petticrew, M., Poland, B., Ridde, V., Shoveller, J., Viehbeck, S. and Wight, D. (2018). Taking account of context in population health intervention research: guidance for producers, users and funders of research. https://doi.org/10.3310/cihr-nihr-01

*Cullen, O., & Walsh, C. A. (2020). A Narrative Review of Ethical Issues in Participatory Research with Young People. YOUNG, 28(4), 363-386. https://doi.org/10.1177/1103308819886470

Curran GM, Landes SJ, McBain SA, Pyne JM, Smith JD, Fernandez ME, Chambers DA, Mittman BS (2022). Reflections on 10 years of effectiveness-implementation hybrid studies. *Front Health Serv*, 8;2:1053496. https://doi.org10.3389/frhs.2022.1053496

Curran, G.M., Bauer, M., Mittman, B., Pyne, J.M. and Stetler, C. (2012). Effectiveness-implementation Hybrid Designs. *Medical Care*, 50(3), pp.217–226. https://doi.org/10.1097/mlr.0b013e3182408812

Czosnek, L., Zopf, E.M., Cormie, P., Rosenbaum, S., Richards, J. and Rankin, N.M. (2022). Developing an implementation research logic model: using a multiple case study design to

establish a worked exemplar. *Implementation Science Communications*, 3(1). https://doi.org/10.1186/s43058-022-00337-8

*Damschroder, L.J. (2020) Clarity out of chaos: use of theory in implementation research. *Psychiatry Research*, 283:1–6. https://doi.org/10.1016/j.psychres.2019.06.036

Damschroder, L.J., Aron, D.C., Keith, R.E., Kirsh, S.R., Alexander, J.A. and Lowery, J.C. (2009). Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implementation Science*, 4(1).

Damschroder, L.J., Reardon, C.M., Opra Widerquist, M.A. and Lowery, J. (2022a). Conceptualizing outcomes for use with the Consolidated Framework for Implementation Research (CFIR): the CFIR Outcomes Addendum. *Implementation Science*, 17(1). https://doi.org/10.1186/s13012-021-01181-5

Damschroder, L.J., Reardon, C.M., Widerquist, M.A.O. and Lowery, J. (2022b). The updated Consolidated Framework for Implementation Research based on user feedback. *Implementation Science*, 17(1), pp.1–16. https://doi.org/10.1186/s13012-022-01245-0

*Dalkin, S. M., Greenhalgh, J., Jones, D., Cunningham, B., & Lhussier, M. (2015). What's in a mechanism? Development of a key concept in realist evaluation. *Implementation Science 10*: 49.

Davies, P., Walker, A.E. and Grimshaw, J.M. (2010). A systematic review of the use of theory in the design of guideline dissemination and implementation strategies and interpretation of the results of rigorous evaluations. *Implementation Science*, 5(1). https://doi.org/10.1186/1748-5908-5-14

De Silva, M.J., Breuer, E., Lee, L., Asher, L., Chowdhary, N., Lund, C. and Patel, V. (2014). Theory of Change: a theory-driven approach to enhance the Medical Research Council's framework for complex interventions. *Trials*, 15(1). https://doi.org/10.1186/1745-6215-15-267

Durlak, J.A. and DuPre, E.P. (2008). Implementation Matters: A Review of Research on the Influence of Implementation on Program Outcomes and the Factors Affecting Implementation. *American Journal of Community Psychology*, 41(3-4), pp.327–350 https://doi.org/10.1007/s10464-008-9165-0

*Eccles, M., Grimshaw, J., Walker, A., Johnston, M., Pitts, N. (2005) Changing the behavior of healthcare professionals: the use of theory in promoting the uptake of research findings. *Journal of Clinical Epidemiology*, 58:107–12.

Edmunds, S.R., Frost, K.M., Sheldrick, R.C., Bravo, A., Straiton, D., Pickard, K., Grim, V., Drahota, A., Kuhn, J., Azad, G., Pomales Ramos, A., Ingersoll, B., Wainer, A., Ibanez, L.V., Stone, W.L., Carter, A. and Broder-Fingert, S. (2022). A method for defining the CORE of a psychosocial intervention to guide adaptation in practice: Reciprocal imitation teaching as a case example. *Autism*. https://doi.org/10.1177/13623613211064431

*Ehrhart, M.G., Aarons, G.A. & Farahnak, L.R. (2014). Assessing the organizational context for EBP implementation: the development and validity testing of the Implementation Climate Scale (ICS). *Implementation Science,* 9, 157. https://doi.org/10.1186/s13012-014-0157-1

Fletcher, A., Jamal, F., Moore, G., Evans, R.E., Murphy, S. and Bonell, C. (2016). Realist complex intervention science: Applying realist principles across all phases of the Medical Research Council framework for developing and evaluating complex interventions. *Evaluation*, 22(3), pp.286–303. https://doi.org/10.1177/1356389016652743

Frye, A.W. and Hemmer, P.A. (2012). Program evaluation models and related theories: AMEE Guide No. 67. *Medical Teacher*, 34(5),e288–e299. https://doi.org/10.3109/0142159x.2012.668637

*Glasgow, R.E., Harden, S.M., Gaglio, B., Rabin, B., Smith, M.L., Porter, G.C., Ory, M.G., & Estabrooks, P.A. (2019). RE-AIM Planning and Evaluation Framework: Adapting to New Science and Practice With a 20-Year Review. *Frontiers in Public Health*, *7*, 64. https://doi.org/10.3389/fpubh.2019.00064

*Graham, I.D., Logan, J., Harrison, M.B., Straus, S.E., Tetroe, J., Caswell, W., Robinson, N. (2006). Lost in knowledge translation: time for a map? *Journal of Continuing Education in Health Professions*, 26:13–24.

Greenhalgh, J. and Manzano, A. (2021). Understanding 'context' in realist evaluation and synthesis. *International Journal of Social Research Methodology*, 25(5), pp.1–14. https://doi.org/10.1080/13645579.2021.1918484

*Greenhalgh, T., Robert, G., Macfarlane, F., et al. (2004). *Diffusion of innovations in service organizations: Systematic review and recommendations*. *The Milbank Quarterly, 82*(4), 581-629. https://doi.org/10.1111/j.0887-378X.2004.00325

*Griffiths, S., Weston, L., Morgan-Trimmer, S., Wheat, H., Gude, A., Manger, L., Oh, T. M., Clarkson, P., Quinn, C., Sheaff, R., Clark, M., Sherriff, I., & Byng, R. (2022). Engaging Stakeholders in Realist Programme Theory Building: Insights from the Prospective Phase of a Primary Care Dementia Support Study. International Journal of Qualitative Methods, 21. https://doi.org/10.1177/16094069221077521

Gustafson, P., Yasmin Abdul Aziz, Lambert, M., Bartholomew, K., Rankin, N., Fusheini, A., Brown, R., Carswell, P., Mihi Ratima, Priest, P. and Crengle, S. (2023). A scoping review of equity-focused implementation theories, models and frameworks in healthcare and their application in addressing ethnicity-related health inequities. *Implementation Science*, 18(1). https://doi.org/10.1186/s13012-023-01304-0

*Harvey G., & Kitson, A. (2016). PARIHS revisited: from heuristic to integrated frame- work for the successful implementation of knowledge into practice. *Implementation Science*, 11:33.

*Harvey G, Fitzgerald L, Fielden S, McBride A, Waterman H, Bamford D, Kislov R, Boaden R. (2011). The NIHR Collaboration for Leadership in Applied Health Research and Care (CLAHRC) for Greater Manchester: combining empirical, theoretical and experiential evidence to design and evaluate a large-scale implementation strategy. *Implementation Science*; 6(96).

*Hedstrom, P., & Swedberg, R. (Eds.). (1998). *Social mechanisms: An analytical approach to social theory*. Cambridge, UK: Cambridge University Press.

Holtrop, J.S., Estabrooks, P.A., Gaglio, B., Harden, S.M., Kessler, R.S., King, D.K., Kwan, B.M., Ory, M.G., Rabin, B.A., Shelton, R.C. and Glasgow, R.E. (2021). Understanding and applying the RE-AIM framework: Clarifications and resources. *Journal of Clinical and Translational Science*, [online] 5(1). https://doi.org/10.1017/cts.2021.789

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. and Kerr, K. (2016a). *Implementation and process evaluations (IPE) for interventions in education settings: A synthesis of the literature*. [online]. Available at: https://research.manchester.ac.uk/en/publications/implementation-and-process-evaluations-ipe-for-interventions-in-e-2

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K. and Kerr, K. (2016b). Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R., & Kerr, K. (2016). Implementation and process evaluation (IPE) for interventions in educational settings: An introductory handbook. [online] Available at: https://www.researchgate.net/publication/328761292_Humphrey_N_Lendrum_A_Ashworth_E_Frearson_K_Buck_R_Kerr_K_2016_Implementation_and_process_evaluation_IPE_for_interventions_in_educational_settings_An_introductory_handbook

Kamphuis, C.B.M., Forberger, S., Lien, N., Rehfuess, E. and Luszczynska, A. (2022). Development of a framework to guide research into policies promoting physical activity and healthy diets in the European context: the system-based Policy Evaluation Network (PEN)

framework. *European Journal of Public Health*, [online] 32(Supplement_4), pp.iv3–iv9. https://doi.org/10.1093/eurpub/ckac068

*Kislov R, Pope C, Martin GP, Wilson PM (2019). Harnessing the power of theorising in implementation science. *Implementation Science*, 11;14(1):103. https://doi.org/10.1186/s13012-019-0957-4

Lafferty, C.K. and Mahoney, C.A. (2003). A Framework for Evaluating Comprehensive Community Initiatives. *Health Promotion Practice*, 4(1), pp.31–44. https://doi.org/10.1177/1524839902238289

Lemire, C., Rousseau, M. and Dionne, C. (2022). A Comparison of Fidelity Implementation Frameworks Used in the Field of Early Intervention. *American Journal of Evaluation*, p.109821402110089. https://doi.org/10.1177/10982140211008978

Lemire, S., Kwako, A., Nielsen, S.B., Christie, C.A., Donaldson, S.I. and Leeuw, F.L. (2020). What Is This Thing Called a Mechanism? Findings From a Review of Realist Evaluations. *New Directions for Evaluation*, 2020(167), pp.73–86. https://doi.org/10.1002/ev.20428

Lengnick-Hall, R., Williams, N.J., Ehrhart, M.G., Willging, C.E., Bunger, A.C., Beidas, R.S. and Aarons, G.A. (2023). Eight characteristics of rigorous multilevel implementation research: a step-by-step guide. *Implementation Science*, 18(1). https://doi.org/10.1186/s13012-023-01302-2

*Logan, J., Graham, I. (2010) *The Ottawa Model of Research Use*. In: Rycroft-Malone J, Bucknall T, editors. *Models and Frameworks for Implementing Evidence-Based Practice: Linking Evidence to Action*. Oxford: Wiley-Blackwell; p. 83–108.

*Manzano (2016) The Craft of Interviewing in Realist Evaluation. *Evaluation*, 1-19. https://doi.org/10.1177/1356389016638615evi.sagepub.com

Masterson-Algar, P., Burton, C.R. and Rycroft-Malone, J. (2018). The generation of consensus guidelines for carrying out process evaluations in rehabilitation research. *BMC Medical Research Methodology*, 18(1). https://doi.org/10.1186/s12874-018-0647-y

May, C., & Finch, T. (2009). Implementing, embedding, and integrating practices: an outline of normalization process theory. *Sociology*, *43*(3), 535-554.

*Mayne, J. (2011). Contribution analysis: Addressing cause and effect. *Evaluating the complex*, pp.53-96.

*Mayne, J. (2019). Revisiting contribution analysis. *Canadian Journal of Program Evaluation*, *34*(2), pp.171-191.

*Mettert, K., Lewis, C., Dorsey, C., Halko, H., & Weiner, B. (2020). Measuring implementation outcomes: An updated systematic review of measures' psychometric properties. Implementation Research and Practice, 1. https://doi.org/10.1177/2633489520936644

* Michie S, Johnston M, Abraham C, Lawton R, Parker D, Walker A. (2005) Making psychological theory useful for implementing evidence based practice: a consensus approach. *Quality and Safety in Health Care*, 14(1):26–33. https://doi.org/10.1136/qshc.2004.011155

*Michie, S., Fixsen, D., Grimshaw, J., & Eccles, M. (2009). Specifying and reporting complex behavior change interventions: The need for a scientific method. *Implementation Science*, 4(1), 40.

*Minkler, M., Salvatore, A.L., & Chang, C. (2017). Participatory approaches for study design and analysis in dissemination and implementation research. In: *Dissemination and implementation research in health: Translating science to practice (2nd Edition)*. Edited by Brownson RC, Colditz GA, Proctor EK. New York: Oxford University Press, pp.175-190.

Moore, G.F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., Moore, L., O'Cathain, A., Tinati, T., Wight, D. and Baird, J. (2015). Process evaluation of complex interventions: Medical Research Council guidance. *BMJ*, [online] 350, pp.h1258–h1258. https://doi.org/10.1136/bmj.h1258

Moullin, J.C., Dickson, K.S., Stadnick, N.A., Albers, B., Nilsen, P., Broder-Fingert, S., Mukasa, B. and Aarons, G.A. (2020). Ten recommendations for using implementation frameworks in research and practice. *Implementation Science Communications*, [online] 1(1). Available at: https://implementationsciencecomms.biomedcentral.com/articles/10.1186/s43058-020-00023-7.

*Moullin JC, Dickson KS, Stadnick NA, Rabin B, Aarons GA. Systematic review of the Exploration, Preparation, Implementation, Sustainment (EPIS) framework. Implement Sci. 2019 Jan 5;14(1):1. https://doi.org/10.1186/s13012-018-0842-6

*Murphy, S.A. (2005) An experimental design for the development of adaptive treatment strategies. *Statistics in Medicine,* 24: 1455–1481.

Nielsen, K. and Abildgaard, J.S. (2013). Organizational interventions: A research-based framework for the evaluation of both process and effects. *Work & Stress*, 27(3), pp.278–297. https://doi.org/10.1080/02678373.2013.812358

Nielsen, K., De Angelis, M., Innstrand, S.T. and Mazzetti, G. (2022). Quantitative process measures in interventions to improve employees' mental health: A systematic literature review and the IPEF framework. *Work & Stress*, 37(1), pp.1–26. https://doi.org/10.1080/02678373.2022.2080775

*Nilsen, P. (2024a) *'Theories and concepts from other fields of potential utility for implementation since'* in Nilsen P (ed) Implementation Science Theory and Application. Routledge. London.

*Nilsen, P. (2024b) 'Determinant frameworks' in Nilsen P (ed) Implementation Science Theory and Application. Routledge. London.

Nilsen, P. (ed) (2024c) Implementation Science Theory and Application*. Routledge. London.

Nilsen, P. (2015). Making Sense of Implementation Theories, Models and Frameworks. *Implementation Science*, 10(53). https://doi.org/10.1186/s13012-015-0242-0

*Nilsen, P., and Moore., J. (2024) 'Process models' in Nilsen, P. (ed) Implementation Science Theory and Application. Routledge. London

Ogden, T. and Fixsen, D.L. (2014). Implementation Science: A Brief Overview and Look Ahead. *Zeitschrift für Psychologie*, 222(1), pp.4–11. https://doi.org/10.1027/2151-2604/a000160

*Palmer, R.C., Ismond, D., Rodriquez, E.J., Kaufman, J.S. (2019) Social determinants of health: future directions for health disparities research. *American Journal of Public Health*, 109(S1):S70–1. https://doi.org/10.2105/AJPH.2019.304964

*Pawson, R., & Tilley, N. (2004). *Realist evaluation*. Accessible from: http://www.communitymatters.com.au/RE_chapter.pdf

*Pawson, R., & Tilley, N. (2008). *Realistic evaluation*. Thousand Oaks, CA: Sage.

*Pawson, R. and Manzano-Santaella, A. (2012). A Realist Diagnostic Workshop. *Evaluation* [online] 18(2), pp.176–191. https://doi.org/10.1177/1356389012440912

Pfadenhauer, L.M., Gerhardus, A., Mozygemba, K., Lysdahl, K.B., Booth, A., Hofmann, B., Wahlster, P., Polus, S., Burns, J., Brereton, L. and Rehfuess, E. (2017). Making sense of

complexity in context and implementation: the Context and Implementation of Complex Interventions (CICI) framework. *Implementation Science*, 12(1). https://doi.org/10.1186/s13012-017-0552-5

*Powell, B.J., Waltz, T.J., Chinman, M.J. Damschroder, L.J., Smith, J.L., Matthieu, M.M., Proctor, E.K. and Kirchner, J.E. (2015). A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. *Implementation Science*, 10, 21. https://doi.org/10.1186/s13012-015-0209-1

Presseau, J., McCleary, N., Lorencatto, F., Patey, A.M., Grimshaw, J.M. and Francis, J.J. (2019). Action, actor, context, target, time (AACTT): a framework for specifying behaviour. *Implementation Science*, 14(1). https://doi.org/10.1186/s13012-019-0951-x

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R. and Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research*, 38(2), pp.65–76. https://doi.org/10.1007/s10488-010-0319-7

Proctor, E.K., Bunger, A.C., Lengnick-Hall, R., Gerke, D.R., Martin, J., Phillips, R. and Swanson, J.C. (2023). Ten years of implementation outcomes research: a scoping review. *Implementation Science*, 18(1). https://doi.org/10.1186/s13012-023-01286-z

Renmans, D. and Castellano Pleguezuelo, V. (2023). Methods in realist evaluation: A mapping review. *Evaluation and Program Planning*, 97, p.102209. https://doi.org/10.1016/j.evalprogplan.2022.102209

Ringhofer, L. and Kohlweg, K. (2019). Has the Theory of Change established itself as the better alternative to the Logical Framework Approach in development cooperation programmes? *Progress in Development Studies*, 19(2), pp.112–122. https://doi.org/10.1177/1464993418822882

*Rogers, E.M. (2003). *Diffusion of Innovations*. 5th ed. New York: Free Press.

*Rolfe, S. (2019). Combining Theories of Change and Realist Evaluation in Practice: Lessons from a Research on Evaluation Study. *Evaluation* [online] 25(3), p.135638901983522. https://doi.org/10.1177/1356389019835229.\

Romão, D.M.M., Setti, C., Arruda, L.H.M., de Melo, R.C., de Araujo, B.C., Tan, A.R., DeMaio, P.N. and Kuchenmüller, T. (2023). Integration of evidence into Theory of Change frameworks in the healthcare sector: A rapid systematic review. *PLOS ONE*, 18(3), p.e0282808. https://doi.org/10.1371/journal.pone.0282808

*Rowland, J., Wills, E., Ott, E. (2024). *Youth Participatory Research: A Review of Reviews and Practice Guidance.* Centre for Evidence and Implementation.

*Saldana L, Schaper H, Campbell M, Chapman J. Standardized Measurement of Implementation: The Universal SIC. Implement Sci. 2015 Aug 20;10(Suppl 1):A73. https://doi.org/10.1186/1748-5908-10-S1-A73

*Scheirer, M.A. (2005). Is sustainability possible? A review and commentary on empirical studies of program sustainability. *American Journal of Evaluation*, 26(3), 320-347.

Sekhon, M., Cartwright, M. and Francis, J.J. (2017). Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. *BMC Health Services Research*, 17(1). https://doi.org/10.1186/s12913-017-2031-8

Seward, N., Hanlon, C., Hinrichs-Kraples, S., Lund, C., Murdoch, J., Taylor Salisbury, T., Verhey, R., Shidhaye, R., Thornicroft, G., Araya, R. and Sevdalis, N. (2021). A guide to systems-level, participatory, theory-informed implementation research in global health. *BMJ Global Health*, 6(12), p.e005365. https://doi.org/10.1136/bmjgh-2021-005365

Skivington, K., Matthews, L., Simpson, S.A., Craig, P., Baird, J., Blazeby, J.M., Boyd, K.A., Craig, N., French, D.P., McIntosh, E., Petticrew, M., Rycroft-Malone, J., White, M. and Moore, L. (2021). A New Framework for Developing and Evaluating Complex interventions: Update of Medical Research Council Guidance. *BMJ*, [online] 374(1), p.n2061. https://doi.org/10.1136/bmj.n2061

*Smith, S.N., Almirall, D., Choi, S.Y., Koschmann, E., Rusch, A., Bilek, E., Lane, A., Abelson, J.L., Eisenberg, D., Himle, J.A., Fitzgerald, K.D., Liebrecht, C., Kilbourne, A.M. (2022). Primary aim results of a clustered SMART for developing a school-level, adaptive implementation strategy to support CBT delivery at high schools in Michigan. *Implementation Science: IS*, *17*(1), 54. https://doi.org/10.1186/s13012-022-01229-0

*Smith, J., Li, D., & Rafferty, M. (2020). The implementation research logic model: a method for planning, executing, reporting, and synthesizing implementation projects. https://doi.org/10.1101/2020.04.05.20054379

*Stirman, S.W., Miller, C.J., Toder, K., Calloway, A. (2013) Development of a framework and coding system for modifications and adaptations of evidence-based interventions. *Implementation Science*, 8:65.

*Suchman, E. (1967). *Evaluative Research: Principles and Practice in Public Service and Social Action Programs.* New York: Russell Sage Foundation.

*Thomas K and Nilsen P (2024) 'Implementation theories' in Nilsen P (ed) Implementation Science Theory and Application. Routledge. London

Vicki, B., Huong, T., Miranda, B., Rachel, L. and Marj, M. (2020). A narrative review of economic constructs in commonly used implementation and scale-up theories, frameworks and models. *Health Research Policy and Systems*, 18(1). https://doi.org/10.1186/s12961-020-00633-6

Waltz, T.J., Powell, B.J., Matthieu, M.M., Damschroder, L.J., Chinman, M.J., Smith, J.L., Proctor, E.K. and Kirchner, J.E. (2015). Use of concept mapping to characterize relationships among implementation strategies and assess their feasibility and importance: results from the Expert Recommendations for Implementing Change (ERIC) study. *Implementation Science*, 10(1). https://doi.org/10.1186/s13012-015-0295-0

Wang, Y., Wong, E.L.-Y., Nilsen, P., Chung, V.C.-H., Tian, Y. and Yeoh, E.-K. (2023). A scoping review of implementation science theories, models, and frameworks - an appraisal of purpose, characteristics, usability, applicability, and testability. *Implementation science: IS*, [online] 18(1), p.43. https://doi.org/10.1186/s13012-023-01296-x

Weiner, B.J., Lewis, C.C., Stanick, C., Powell, B.J., Dorsey, C.N., Clary, A.S., Boynton, M.H. and Halko, H. (2017). Psychometric assessment of three newly developed implementation outcome measures. *Implementation Science*, [online] 12(1). https://doi.org/10.1186/s13012-017-0635-3

*Weiner, B.J., Mettert, K.D., Dorsey, C.N., Nolen, E.A., Stanick, C., Powell, B.J., & Lewis, C.C. (2020). Measuring readiness for implementation: A systematic review of measures' psychometric and pragmatic properties. *Implementation Research and Practice*, *1*, 2633489520933896. https://doi.org/10.1177/2633489520933896

*Weiss, C.H. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families. *New approaches to evaluating community initiatives: Concepts, methods, and contexts*, *1*, 65-92.

*Weiss, C.H. (1997). Theory-based evaluation: past, present, and future. *New Directions for Evaluation*, *76*, 41-55.

*Wensing M. (2021). Reflections on the measurement of implementation constructs. *Implementation Research and Practice*, 11; 2:26334895211020125. https://doi.org/10.1177/26334895211020125

Wolfenden, L., Foy, R., Presseau, J., Grimshaw, J.M., Ivers, N.M., Powell, B.J., Taljaard, M., Wiggers, J., Sutherland, R., Nathan, N., Williams, C.M., Kingsland, M., Milat, A., Hodder, R.K. and Yoong, S.L. (2021). Designing and undertaking randomised implementation trials: guide for researchers. *BMJ*, 372(372), p.m3721. https://doi.org/10.1136/bmj.m3721

Woodland, R.H. and Mazur, R. (2024). Reclaiming Logic Modeling for Evaluation: A Theory of Action Framework. *American journal of evaluation/˜The American journal of evaluation*. https://doi.org/10.1177/10982140231224495

Woodward, E.N., Singh, R.S., Ndebele-Ngwenya, P., Melgar Castillo, A., Dickson, K.S. and Kirchner, J.E. (2021). A more practical guide to incorporating health equity domains in implementation determinant frameworks. *Implementation Science Communications*, 2(1). https://doi.org/10.1186/s43058-021-00146-5